CHAIN: Journal of Computer Technology, Computer Engineering and Informatics
Website : https://ejournal.techcart-press.com/index.php/chain
E-ISSN 2964-2485
P-ISSN 2964-2450
Volume 4, Number 3, July 2026
DOI : **https://doi.org/10.58602/chain.v4i3.281**
Page 229-250

Klasterisasi Toko Berdasarkan Monitoring Summary Audit Kepuasan Pelanggan Menggunakan Algoritma BIRCH

Maulana Agung Saputro[1*], Puspita Maelani[2], Rifka Audinasari[3]
[1,2]Akuntansi, Politeknik Negeri Jakarta, Indonesia
[3]Keuangan dan Perbankan Terapan, Politeknik Negeri Jakarta, Indonesia
[1*]maulana.agungsaputro@lecturer.pnj.ac.id, [2] puspita.maelani@akuntansi.pnj.ac.id,
[3]rifka.audinasari@akuntansi.pnj.ac.id

Abstrak: Audit kepuasan pelanggan internal merupakan fungsi penilaian independen yang penting bagi perusahaan ritel untuk menjaga konsistensi kualitas layanan di seluruh jaringan toko. PT XYZ memiliki 210 toko di Indonesia dan secara rutin melaksanakan audit berbasis web. Namun, Supervisor Toko masih mengalami kesulitan dalam menentukan prioritas pelatihan dan evaluasi karena belum mengetahui parameter checklist mana yang paling berpengaruh terhadap penurunan final score toko. Penelitian ini bertujuan membangun model klasterisasi toko berdasarkan karakteristik parameter checklist bernilai rendah dengan algoritma BIRCH. Dataset penelitian berjumlah 109.566 data audit tahun 2025 yang berasal dari sistem internal perusahaan. Metode penelitian mengikuti kerangka CRISP-DM yang meliputi Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment. Tahap preprocessing mencakup data cleaning, data selection, data transformation, serta normalisasi Z-Score agar data siap diproses. Proses modeling menggunakan CF-Tree dengan parameter Branching Factor B dan Threshold T yang dimodifikasi secara dinamis untuk meningkatkan kualitas cluster. Evaluasi cluster dilakukan menggunakan Silhouette Coefficient. Hasil penelitian menunjukkan terbentuk 2 cluster optimal dari CF4 dan CF5 dengan nilai SC Cluster 1 sebesar 0,9994, SC Cluster 2 sebesar 0,9988, dan SC Global sebesar 0,9989. Hasil ini menunjukkan struktur cluster yang sangat kuat dan valid. Sistem berbasis Python juga berhasil menampilkan visualisasi scatter plot sebagai dasar pengambilan keputusan evaluasi kinerja karyawan.
Kata Kunci: Audit Kepuasan Pelanggan; Kualitas Pelayanan; BIRCH Clustering; Silhouette Coefficient; Big Data

Abstract: Internal customer satisfaction audit is an important independent assessment function for retail companies to maintain consistent service quality across the store network. PT XYZ has 210 stores in Indonesia and routinely conducts web-based audits. However, Store Supervisors still have difficulty determining training and evaluation priorities because they do not yet know which checklist parameters have the greatest impact on lowering the store's final score. This study aims to build a store clustering model based on the characteristics of low-value checklist parameters using the BIRCH algorithm. The research dataset consists of 109,566 audit records from 2025 obtained

Maulana Agung Saputro: *Penulis Korespondensi

CHAIN: Journal of Computer Technology, Computer Engineering and Informatics
Website : https://ejournal.techcart-press.com/index.php/chain
E-ISSN 2964-2485
P-ISSN 2964-2450
Volume 4, Number 3, July 2026
DOI : **https://doi.org/10.58602/chain.v4i3.281**
Page 229-250

from the company's internal system. The research method follows the CRISP-DM framework, which includes Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The preprocessing stage includes data cleaning, data selection, data transformation, and Z-Score normalization to prepare the data for processing. The modeling process uses a CF-Tree with Branching Factor B and dynamically modified Threshold T parameters to improve cluster quality. Cluster evaluation is carried out using the Silhouette Coefficient. The results show that 2 optimal clusters were formed from CF4 and CF5, with Cluster 1 SC = 0.9994, Cluster 2 SC = 0.9988, and Global SC = 0.9989. These results indicate a very strong and valid cluster structure. The Python-based system also successfully presents a scatter plot visualization as a basis for employee performance evaluation decisions.
Keywords: Customer Satisfaction Audit; Service Quality; BIRCH Clustering; Silhouette Coefficient; Big Data

1. PENDAHULUAN

Kepuasan pelanggan merupakan salah satu indikator utama keberhasilan operasional toko dalam industri ritel modern. Dalam konteks pengelolaan jaringan toko yang berskala besar, monitoring dan evaluasi kepuasan pelanggan secara berkala menjadi krusial untuk memastikan standar layanan yang konsisten di seluruh cabang. Data hasil audit kepuasan pelanggan yang dikumpulkan secara rutin menghasilkan volume data yang besar dan heterogen, sehingga diperlukan pendekatan analisis yang efektif dan efisien untuk mengekstrak pola bermakna darinya (Das et al., 2026).

Klasterisasi atau pengelompokan data merupakan salah satu teknik pembelajaran mesin tanpa pengawasan (unsupervised learning) yang digunakan secara luas untuk menemukan struktur tersembunyi dalam data. Metode ini bekerja dengan mengelompokkan objek-objek yang memiliki kemiripan ke dalam satu klaster, sehingga objek dalam klaster yang sama lebih mirip satu sama lain dibandingkan dengan objek di klaster lain (Głuszczak & Powroźnik, 2026). Dalam konteks bisnis ritel, klasterisasi memungkinkan identifikasi segmentasi toko berdasarkan profil kinerja auditnya, yang selanjutnya dapat digunakan sebagai dasar pengambilan keputusan strategis (Li et al., 2026)

Berbagai algoritma klasterisasi telah dikembangkan dan diterapkan dalam berbagai domain, mulai dari segmentasi pengguna e-commerce (Li et al., 2026), analisis data IoT real-time, hingga pengelompokan data penjualan (Laurenso et al., 2024). Di antara algoritma-algoritma tersebut, BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) menonjol sebagai solusi yang sangat efisien untuk data berskala besar. BIRCH memproses data secara inkremental menggunakan struktur pohon CF (Clustering Feature) yang memungkinkan algoritma ini bekerja dalam satu kali pemindaian data dengan kebutuhan memori yang terkontrol (Hasan, 2024).

Algoritma clustering berbasis kepadatan yang banyak digunakan dalam proses data mining untuk menemukan pola pada data berukuran besar. Algoritma ini mengelompokkan objek berdasarkan tingkat kepadatan data sehingga mampu membentuk cluster dengan bentuk yang tidak beraturan serta mengidentifikasi data pencilan (noise) secara otomatis (Zhang, 2022).

Penelitian sebelumnya telah membuktikan efektivitas BIRCH dalam berbagai konteks. Das et al. (Das et al., 2026) mengembangkan kerangka kerja cerdas berbasis BIRCH yang tangguh untuk klasterisasi data IoT real-time, menunjukkan bahwa BIRCH mampu menangani aliran data kontinu dengan latensi rendah, mengusulkan BIRCH-AE, sebuah kerangka ensemble hierarkis yang menggabungkan BIRCH dengan autoencoder untuk segmentasi pengguna e-commerce berskala besar, dan berhasil meningkatkan kualitas klaster secara signifikan (Li et al., 2026). Sementara itu, Laurenso et al. (Laurenso et al.,

Maulana Agung Saputro: *Penulis Korespondensi

CHAIN: Journal of Computer Technology, Computer Engineering and Informatics
Website : https://ejournal.techcart-press.com/index.php/chain
E-ISSN 2964-2485
P-ISSN 2964-2450
Volume 4, Number 3, July 2026
DOI : **https://doi.org/10.58602/chain.v4i3.281**
Page 229-250

2024) membandingkan K-Means, Hierarchical Clustering, dan BIRCH untuk menentukan target pemasaran penjualan vape di Indonesia, dengan BIRCH menunjukkan keunggulan dari sisi skalabilitas. Rahmadani (Sari et al., 2024) juga menerapkan BIRCH untuk mengelompokkan tingkat penjualan smartphone pada toko offline, membuktikan kemampuan algoritma ini dalam konteks ritel Indonesia.

Salah satu parameter penting dalam BIRCH adalah radius ambang batas (threshold radius). Hasan(Hasan, 2024) dalam penelitiannya secara khusus menganalisis dampak radius pada algoritma BIRCH terhadap distribusi dan kualitas klaster, menyimpulkan bahwa pemilihan radius yang tepat sangat menentukan kualitas hasil pengelompokan. Temuan ini relevan dengan penelitian yang dilakukan Das (Das et al., 2026) dan BIRCH-AE yang juga menekankan pentingnya konfigurasi parameter dalam mencapai hasil klasterisasi yang optimal (Li et al., 2026).

Evaluasi kualitas klaster merupakan tahapan kritis dalam proses klasterisasi. Vardakas et al. memperkenalkan metrik soft silhouette score untuk klasterisasi mendalam yang bertujuan menghasilkan klaster yang kompak dan terpisah dengan baik (Vardakas et al., 2024). Pendekatan evaluasi ini melengkapi metrik klasik seperti Davies-Bouldin Index (DBI) dan Calinski-Harabasz Index yang umum digunakan dalam penerapan BIRCH (Hasan, 2024). Selain itu, Głuszczak dan Powroźnik memberikan tinjauan komprehensif tentang berbagai metode klasterisasi dalam pembelajaran mesin, menekankan bahwa pemilihan algoritma yang tepat harus disesuaikan dengan karakteristik data dan tujuan analisis (Głuszczak & Powroźnik, 2026).

Pada penelitian ini dibutuhkan metode Bayesian Z-score yang dikembangkan untuk menghasilkan penilaian yang lebih reliabel terhadap pengukuran diameter aorta dengan mempertimbangkan ketidakpastian data dan distribusi probabilistik. Pendekatan ini memungkinkan interpretasi nilai Z-score yang lebih akurat dibandingkan metode konvensional, sehingga dapat meningkatkan kualitas analisis data yang membutuhkan proses standardisasi dan identifikasi penyimpangan terhadap nilai rata-rata populasi. Dalam analisis statistik, Z-score berfungsi untuk menunjukkan seberapa jauh suatu nilai menyimpang dari rata-rata dalam satuan standar deviasi, sehingga memudahkan proses deteksi outlier dan perbandingan antar data yang memiliki skala berbeda (Bindini et al., 2026).

Dari sisi metodologi penelitian, Mendes dan Farinha mengusulkan kerangka MIDA (Method for Industrial Data Analysis) yang berbasis CRISP-DM sebagai panduan terstruktur untuk analisis data industry (Mendes & Farinha, 2026). Kerangka ini menekankan pentingnya tahapan pra-pemrosesan data, pemilihan algoritma yang tepat, dan validasi hasil yang sistematis—prinsip yang diadopsi dalam penelitian ini. Fujino et al. juga memberikan kontribusi dalam representasi trajektori berbasis distribusi kode untuk penerapan klasterisasi pada data AIS berskala besar, memperkuat pentingnya representasi fitur yang tepat sebelum klasterisasi dilakukan (Fujino et al., 2026).

Penelitian ini bertujuan untuk menerapkan algoritma BIRCH dalam mengklasterisasi toko-toko berdasarkan data monitoring summary audit kepuasan pelanggan. Dengan mengidentifikasi kelompok toko yang memiliki profil kinerja serupa, manajemen jaringan toko dapat merancang strategi perbaikan yang lebih terarah dan efektif. Kontribusi utama penelitian ini meliputi: (1) penerapan BIRCH pada dataset audit kepuasan pelanggan ritel berskala nyata; (2) analisis pengaruh parameter threshold terhadap kualitas klaster; dan (3) interpretasi hasil klaster sebagai dasar rekomendasi strategis operasional toko.

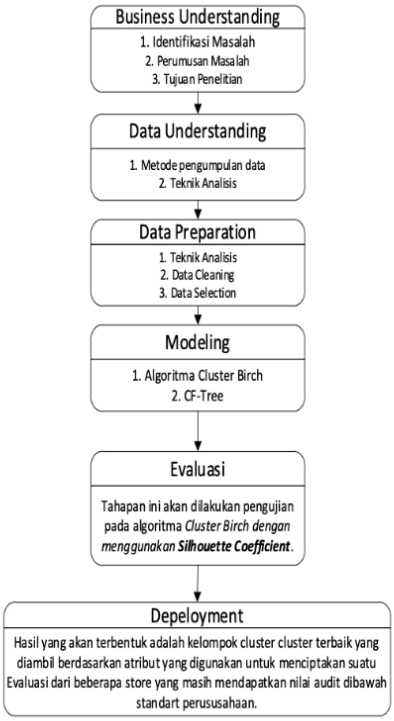
2. METODE PENELITIAN

Desain Penelitian

Maulana Agung Saputro: *Penulis Korespondensi

CHAIN: Journal of Computer Technology, Computer Engineering and Informatics
Website : https://ejournal.techcart-press.com/index.php/chain
E-ISSN 2964-2485
P-ISSN 2964-2450
Volume 4, Number 3, July 2026
DOI : **https://doi.org/10.58602/chain.v4i3.281**
Page 229-250

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining yang mengikuti kerangka CRISP-DM. Tahapan penelitian dilaksanakan secara sistematis mulai dari Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, hingga Deployment. Data yang digunakan adalah data primer berupa hasil ekspor dari sistem web internal audit PT XYZ 1 tahun di 2026 yang berjumlah 109.566 baris data, serta data sekunder berupa literatur dan jurnal ilmiah yang relevan.



Gambar 1. Langkah-langkah penelitian

Business Understanding
Pada tahap ini dilakukan pemahaman mendalam terhadap proses bisnis PT XYZ. SPV Toko menghadapi kesulitan dalam memperbaiki nilai toko setelah dilaksanakan audit internal kepuasan pelanggan karena tidak mengetahui bagian checklist mana yang mendapat nilai rendah. Tujuan bisnis penelitian ini adalah membangun model klasterisasi yang dapat mengelompokkan toko berdasarkan parameter checklist yang nilainya di bawah standar perusahaan, sehingga SPV Toko dapat menentukan program pelatihan dan evaluasi yang tepat sasaran.

Tabel 1. Gap Analisis

Aspek	Harapan	Realita
Model klasterisasi	Terbentuknya model pengelompokan toko berdasarkan parameter checklist yang masih rendah dengan sub-cluster menggunakan algoritma BIRCH	Audit hanya berjalan mengikuti prosedur perusahaan tanpa adanya model klasterisasi yang dapat mengidentifikasi toko bermasalah secara otomatis

Maulana Agung Saputro: *Penulis Korespondensi

CHAIN: Journal of Computer Technology, Computer Engineering and Informatics
Website : https://ejournal.techcart-press.com/index.php/chain
E-ISSN 2964-2485
P-ISSN 2964-2450
Volume 4, Number 3, July 2026
DOI : **https://doi.org/10.58602/chain.v4i3.281**
Page 229-250

Evaluasi SPV	SPV dapat memberikan pelatihan dan evaluasi berbasis data yang spesifik per parameter	SPV kesulitan mengidentifikasi parameter checklist mana yang paling berpengaruh terhadap rendahnya final score

Pemahaman Bisnis (Business Understanding) Pada tahapan ini peneliti memahami informasi mengenai proses bisnis dari objek penelitian. Dimana dari hal tersebut dapat ditemukan masalah dari adanya nilai dashboard monitoring yang masih tidak memenuhi standart kepuasan pelanggan di beberapa Toko, maka dari itu peneliti membuat suatu Evaluasi untuk meningkatkan pelayanan karyawan dengan menggunakan perhitungan algoritma BIRCH untuk mengetahui sub-sub Cluster dari checklist/kategori mana yang nilainya masih di bawah standart. Dalam melakukan analisis masalah, penulis menggunakan tabel gap analysis untuk melihat kondisi saat ini dengan kondisi saat ini. Tabel gap analysis dapat dilihat pada tabel berikut :

Tabel 2. Gap Analisis

Harapan	Terbentuknya sebuah model pengelompokan Toko berdasarkan perhitungan *Monitoring* Summary dalam menentukan nilai karakteristik parameter *checklist* Toko yang masih rendah dengan pembagian *Sub-Cluster* menggunakan algoritma *BIRCH*.
Realita	Audit hanya berjalan mengikuti prosedur perusahaan tanpa ada pelatihan dan evaluasi terhadap Toko-Toko yang karakteristik *checklist*nya masih mendapatkan nilai rendah.
Masalah	*SPV* Toko kesulitan dalam memperbaiki nilai-nilai Toko setelah dilakukannya Audit internal kepuasan pelanggan, karena ketidak tahuan karakteristik *checklist* mana yang mendapatkan nilai rendah dari tiap-tiap Toko.
Tujuan	Tujuan dari penelitian ini untuk membantu *SPV* dalam memberikan pelatihan dan mengevaluasi Toko yang belum mencapai nilai sesuai dengan *standart* yang sudah ditentukan perusahaan.
Pertanyaan Penelitian	Bagaimana aturan *Clustering* atau pengelompokan yang akan terbentuk untuk menentukan nilai karakteristik *checklist* Toko yang masih mendapatkan nilai rendah.

Data Understanding

Dataset yang digunakan diperoleh dari hasil ekspor 1 tahun pada sistem web-based audit internal PT XYZ di tahun 2025. Total data berjumlah 109.566 baris dengan 14 atribut awal. Dataset mencakup informasi dari 210 toko Samsung Experience Store (SES) yang tersebar di seluruh Indonesia dengan 5 dimensi penilaian utama dan 295 item checklist.

Tabel 3. Atribut Dataset Awal

No	Atribut	Keterangan
1	Year	Tahun audit dilaksanakan
2	Month	Bulan audit dilaksanakan

Maulana Agung Saputro: *Penulis Korespondensi

CHAIN: Journal of Computer Technology, Computer Engineering and Informatics
Website : https://ejournal.techcart-press.com/index.php/chain
E-ISSN 2964-2485
P-ISSN 2964-2450
Volume 4, Number 3, July 2026
DOI : **https://doi.org/10.58602/chain.v4i3.281**
Page 229-250

3	Name Toko	Nama toko yang diaudit
4	Dimension	Kategori bagian yang dinilai (5 dimensi)
5	Questionnaires	Pertanyaan checklist saat audit
6	Score	Nilai hasil audit per checklist
7	Cdbln	Code atribut Bulan (setelah transformasi)
8	CdToko	Code atribut Toko (setelah transformasi)
9	CdDimensions	Code atribut Dimensions (setelah transformasi)
10	CdQuest	Code atribut Questionnaire (setelah transformasi)
11	Channel Code	ID toko
12	Partner	Perusahaan mitra dalam pelaksanaan audit

Lima dimensi penilaian yang menjadi atribut utama audit adalah:

- **Availability**: kelengkapan stok produk di toko
- **Visibility**: kelengkapan perangkat display dan materi promosi
- **Advocacy**: kerapihan dan kelengkapan seragam staf
- **Customer Experience**: SOP pelayanan dan interaksi dengan pelanggan
- **APS (Accessories Product Service)**: ketersediaan produk aksesori
- **Final Score** : Nilai Akhir

Data Preparation

Tahap Data Preparation meliputi data cleaning (menghapus missing value dan data duplikat), data selection (memilih atribut yang relevan), dan data transformation (mengubah variabel kategorik menjadi numerik). Atribut yang digunakan setelah seleksi adalah: Cd_Bln (kode bulan, X1), Cd_Toko (kode toko, X2), Cd_Quest (kode pertanyaan, X3), dan Cd_Score (kode skor, X4). Normalisasi data dilakukan menggunakan metode Z-Score dengan rumus: newdata = (data - mean) / std. Tahap Data Preparation merupakan proses membersihkan dan mempersiapkan data sebelum dilakukan pemodelan. Data awal memiliki 7 atribut utama yang kemudian direduksi dan ditransformasi. Berikut adalah tabel hasil Data Preparation:

Tabel 4. Atribut Dataset setelah Data Preparation

No	YEAR	CHANNEL CODE	MONTH	NAME TOKO
1	2025	C053401629	Jan	S - AMBON - AMBON CITY CENTER
2	2025	C053401629	Feb	S - AMBON - AMBON CITY CENTER
3	2025	C053401629	Mar	S - AMBON - AMBON CITY CENTER
4	2025	C053401629	Apr	S - AMBON - AMBON CITY CENTER
5	2025	C053401629	Mei	S - AMBON - AMBON CITY CENTER
6	2025	C053401629	Jun	S - AMBON - AMBON CITY CENTER

Maulana Agung Saputro: *Penulis Korespondensi

CHAIN: Journal of Computer Technology, Computer Engineering and Informatics
Website : https://ejournal.techcart-press.com/index.php/chain
E-ISSN 2964-2485
P-ISSN 2964-2450
Volume 4, Number 3, July 2026
DOI : **https://doi.org/10.58602/chain.v4i3.281**
Page 229-250

7	2025	C053401629	Jul	S - AMBON - AMBON CITY CENTER
8	2025	C053401629	Aug	S - AMBON - AMBON CITY CENTER
9	2025	C068211571	Sep	S - AMBON - AMBON CITY CENTER
10	2025	C068211571	Okt	S - AMBON - AMBON CITY CENTER
11	2025	C068211571	Nov	S - AMBON - AMBON CITY CENTER
dst	….	….	….	….
12	2025	C068211571	Des	S - AMBON - AMBON CITY CENTER

Tabel 5. Perubahan Data Atribut (Lanjutan)

DIMENTIONS	***QUESTIONNAIRES***	***Score***
1. Availability	GLX 1	NA
2. Visibility	Display GLX	1
3. Advocacy	Staff	1
4. Customer Experience	Service Staff	1
5. APS	GLX note Prime	1

Data Cleaning

Proses data cleaning dilakukan dengan cara menghapus data yang memiliki missing value dari dataset. Setelah data ditarik dari sistem web-based, data digabungkan dengan data audit bulan sebelumnya kemudian dilakukan identifikasi dan penghapusan nilai yang kosong atau tidak valid.

Tahap pra-pemrosesan data dilakukan untuk memastikan kualitas dan kesiapan data sebelum proses klasterisasi. Langkah-langkah pra-pemrosesan meliputi:

- Pembersihan data: penanganan nilai yang hilang (missing values) dan deteksi serta penanganan data pencilan (outlier).
- Normalisasi fitur: penyekalaan nilai fitur menggunakan metode Min-Max Normalization untuk memastikan seluruh atribut berada dalam rentang nilai yang sebanding, sehingga tidak ada atribut yang mendominasi proses klasterisasi (Głuszczak & Powroźnik, 2026).
- Seleksi fitur: pemilihan variabel-variabel yang paling representatif dan relevan terhadap profil kinerja toko berdasarkan domain knowledge dan analisis korelasi, mengacu pada prinsip representasi fitur yang ditekankan oleh Fujino (Fujino et al., 2026)

Data Selection

Dua atribut direduksi dari dataset pada tahap ini, yaitu atribut Channel Code (karena hanya berisi ID toko yang tidak berkontribusi pada proses clustering) dan atribut Partner (karena tidak memiliki kontribusi langsung pada tujuan penelitian). Setelah reduksi, tersisa 5 atribut utama yang digunakan untuk modeling.

Data Trasformation

Pada tahapan ini, hal yang pertama yang akan dilakukan adalah mengubah data yang tidak konsisten dan akan dilakukan perbaikan data untuk daya yang tidak konsisten.

Tabel 6. Perubahan Data Atribut

MONTH	**Cd_Bln**	**NAMA TOKO**	**Cd_Toko**	***DIMENTIONS***

Maulana Agung Saputro: *Penulis Korespondensi

CHAIN: Journal of Computer Technology, Computer Engineering and Informatics
Website : https://ejournal.techcart-press.com/index.php/chain
E-ISSN 2964-2485
P-ISSN 2964-2450
Volume 4, Number 3, July 2026
DOI : **https://doi.org/10.58602/chain.v4i3.281**
Page 229-250

Jan	44191	S - AMBON - AMBON CITY CENTER	613860	1. Availability
Feb	44192	S - AMBON - AMBON CITY CENTER	613860	1. Availability
Mar	44193	S - AMBON - AMBON CITY CENTER	613860	1. Availability
Apr	44194	S - AMBON - AMBON CITY CENTER	613860	1. Availability
Mei	44195	S - AMBON - AMBON CITY CENTER	613860	1. Availability
dst	….	….	….	….

Tabel 7. Perubahan Data Atribut (Lanjutan)

QUESTIONNAIRES	*Cd_Quest*	*Score*	*Cd_Score*
GLX Choose 1 Sku	102100	NA	30910711
GLX FE - Choose 1 Sku	102105	1	30910711
GLX - Mystic Black	102110	1	30910711
GLX - Mystic Brown	102115	1	700173460
GLX- Mystic White	102120	1	30910711
dst	….	….	….

Karena algoritma BIRCH hanya dapat memproses data numerik, seluruh atribut kategorik dikonversi menjadi representasi numerik menggunakan code book yang telah didiskusikan dengan pakar domain. Transformasi yang dilakukan adalah:

- Atribut **Month** diubah menjadi **X1** (kode bulan numerik)
- Atribut **CdToko** diubah menjadi **X2** (kode toko numerik)
- Atribut **CdQuest** diubah menjadi **X3** (kode pertanyaan numerik)
- Atribut **Score** diubah menjadi **X4** (nilai numerik)

Dataset yang digunakan harus dinormalisasi terlebih dahulu agar tidak terjadi berbagai anomali data dan tidak konsistensinya data. Metode untuk normalisasi yang digunakan adalah Z-Score, dapat dilihat pada tabel Perubahan Data Atribut dibawah ini.

Tabel 8. Table Z-Score

No	X1	X2	X3	X4
1	44197	16072150	102105	30910711
2	44197	25615653	102140	30910711
3	44197	39999314	102185	700173460
4	44197	57134146	102220	30910711
5	44197	47377179	102265	700173460
6	44197	39808753	102310	700173460
7	44197	7824629	102375	30910711
dst	…….	…….	…….	…….
109566	44197	46682653	102410	30910711

Maulana Agung Saputro: *Penulis Korespondensi

CHAIN: Journal of Computer Technology, Computer Engineering and Informatics
Website : https://ejournal.techcart-press.com/index.php/chain
E-ISSN 2964-2485
P-ISSN 2964-2450
Volume 4, Number 3, July 2026
DOI : **https://doi.org/10.58602/chain.v4i3.281**
Page 229-250

109567	44197	65333175	102425	700173460
109568	44197	41785794	102550	700173460

Normalisasi Z-Score

Normalisasi menggunakan metode Z-Score dilakukan untuk mencegah terjadinya anomali data dan inkonsistensi akibat perbedaan skala antar atribut. Rumus Z-Score yang digunakan adalah:

$$newdata = (data - mean) / std$$

Proses normalisasi ini menghasilkan dataset yang siap diproses oleh algoritma BIRCH. Contoh data setelah normalisasi Z-Score.

Tabel 9. Sampel Data Setelah Normalisasi Z-Score

No	X1	X2	X3	X4
1	0,41	0,33	0,40	0,26
2	0,41	0,29	0,40	0,26
3	0,41	0,22	0,40	2,83
4	0,41	0,14	0,40	0,26
5	0,41	0,19	0,40	2,83
6	0,41	0,22	0,40	2,83
7	0,41	0,37	0,40	0,26

Modeling

Pre-Clustering dan Pembentukan CF

Proses pre-clustering diawali dengan mengkonversi setiap data point ke dalam format Clustering Feature (CF). Setiap data diubah menjadi CF = (N, LS, SS). Parameter input yang digunakan dalam pembentukan CF-Tree adalah Branching Factor (B = 2), Threshold (T = 2), dan maksimum entry leaf (L = 2). Dalam penelitian ini, nilai Threshold T

Maulana Agung Saputro: *Penulis Korespondensi

CHAIN: Journal of Computer Technology, Computer Engineering and Informatics
Website : https://ejournal.techcart-press.com/index.php/chain
E-ISSN 2964-2485
P-ISSN 2964-2450
Volume 4, Number 3, July 2026
DOI : **https://doi.org/10.58602/chain.v4i3.281**
Page 229-250

dimodifikasi menjadi dinamis (BIRCH CF-Leaf Modif) untuk memperbaiki kualitas cluster yang dihasilkan.

Algoritma BIRCH (Balance Iterative Reducing and Clustering Using Hierarchies) menggunakan konsep Clustering Feature (CF) sebagai representasi kompak data dalam sebuah cluster. CF didefinisikan sebagai triplet statistic :

CF = (N, LS, SS)

LS = Σ Pωi ∈ N

SS = Σ |Pωi ∈ N|²

Keterangan: N = jumlah data dalam cluster, LS = jumlah koordinat data linear, SS = jumlah koordinat kuadrat. Jarak antar cluster dihitung menggunakan Average Inter-Cluster Distance (D2):

$D2 = \sqrt{[(N_1 \cdot SS_2) + (N_2 \cdot SS_1) - 2 \cdot LS_1 \cdot LS_2]} / (N_1 \cdot N_2)$

Nilai radius cluster dihitung dengan rumus:

$R = \sqrt{[(SS - (LS)^2/n) / n]}$

Tabel 10. Hasil Pembentukan Clustering Feature (CF) dari Data Sampel

No	CF	LS (Linear Sum)	SS (Squared Sum)	R
1	CF1	(0,41; 0,33; 0,40; 0,26)	(0,16; 0,11; 0,16; 0,07)	0,1239
2	CF2	(0,41; 0,29; 0,40; 0,26)	(0,16; 0,08; 0,16; 0,07)	1,8
3	CF3	(0,41; 0,22; 0,40; 2,83)	(0,16; 0,05; 0,16; 7,98)	0,36
4	CF4	(1,23; 0,88; 1,20; 3,35)	(0,48; 0,27; 0,48; 8,12)	2,99
5	CF5	(2,47; 1,90; 2,80; 6,96)	(1,12; 0,56; 1,12; 16,31)	3,32
6	CF6	(0,41; 0,22; 0,40; 2,83)	(0,16; 0,05; 0,16; 7,98)	0,0149
7	CF7	(0,41; 0,37; 0,40; 0,26)	(0,16; 0,14; 0,16; 0,07)	0,0125

Proses pembangunan CF-Tree dimulai dari Leaf Node pertama hingga terbentuk Final CF Node. Berikut adalah visualisasi tahapan pembangunan CF-Tree:

Langkah pertama pada proses Clustering adalah sebagai berikut, SubCluster pertama akan masuk kedalam root dan membentuk di CF Node Non leaf dan masuk ke leaf.

$$CF1 = [1; (0{,}41;0{,}33;0{,}4;0{,}26);((0{,}16;0{,}11;0{,}16;0{,}07)]$$

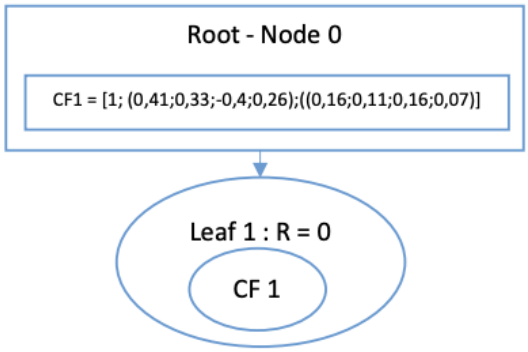
$$R = \frac{\sqrt{0{,}16 - (0{,}19)^2 / 1}}{1}$$

$$R = 0{,}1239$$

Dan membetuk leaf node yang dapat dilihat pada Gambar Leaf Node Pertama:

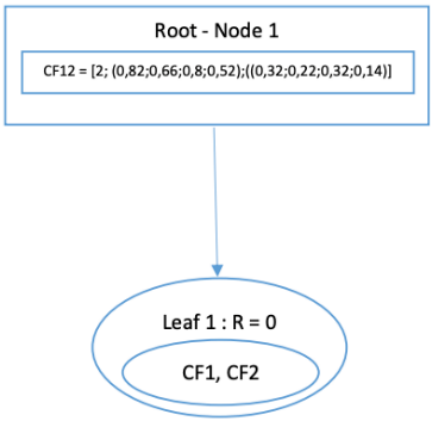
Maulana Agung Saputro: *Penulis Korespondensi

CHAIN: Journal of Computer Technology, Computer Engineering and Informatics
Website : https://ejournal.techcart-press.com/index.php/chain
E-ISSN 2964-2485
P-ISSN 2964-2450
Volume 4, Number 3, July 2026
DOI : **https://doi.org/10.58602/chain.v4i3.281**
Page 229-250



Gambar 2. Leaf Node 1 dan Leaf Node 2 pada CF-Tree

Setelah subCluster CF1 masuk, kemudian subCluster CF2 akan mulai masuk juga kedalam Cluster CF dan akan ada pembaharuan informasi mengenai CF. setelah bergabung dengan CF maka akan di lakukan pengecekkan terhadap radius untuk bergabung dengan CF-Leaf.



Gambar 3. Leaf Node 1 dan Leaf Node 2 pada CF-Tree

$$CF12 = [2; (0{,}82;0{,}66;0{,}8;0{,}52); (0{,}32;0{,}22;0{,}32;0{,}14)]$$

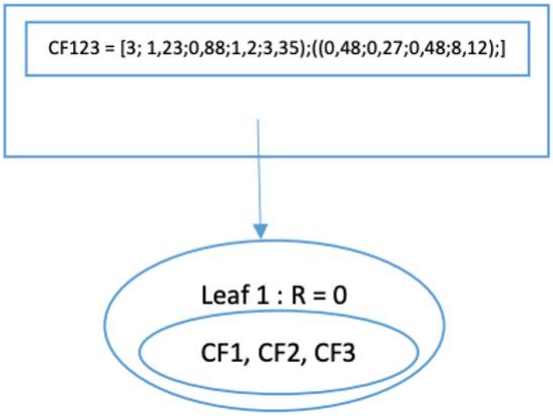
$$R = \frac{\sqrt{2{,}8 \ - (1)^2 \ /2}}{2}$$

$$R = 1{,}8$$

Setelah subCluster CF1 dan CF2 masuk, kemudian subCluster CF3 akan mulai masuk juga kedalam Cluster CF dan akan ada pembaharuan informasi mengenai CF. setelah bergabung dengan CF maka akan di lakukan pengecekkan terhadap radius untuk bergabung dengan CF-Leaf (Nilai R masih tidak melebihi nilai Threshold sehimgga dapat bergabung dengan CFLeaf yang sama dan aka nada pembaruhan informasi mengenai CF).

Maulana Agung Saputro: *Penulis Korespondensi

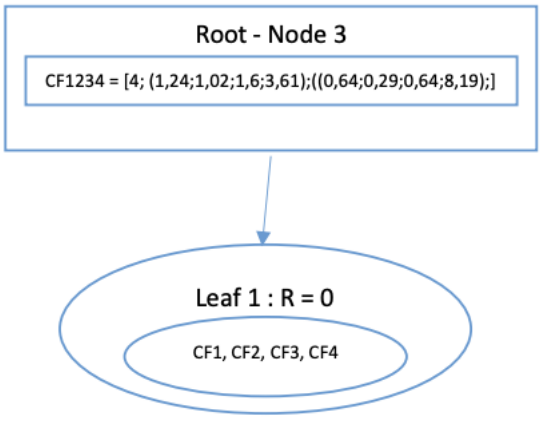
CHAIN: Journal of Computer Technology, Computer Engineering and Informatics
Website : https://ejournal.techcart-press.com/index.php/chain
E-ISSN 2964-2485
P-ISSN 2964-2450
Volume 4, Number 3, July 2026
DOI : **https://doi.org/10.58602/chain.v4i3.281**
Page 229-250



Gambar 4. Leaf Node 3 dan Leaf Node 4 pada CF-Tree

$$CF123 = [3; (1{,}23;0{,}88;1{,}2;3{,}35);(0{,}48;0{,}27;0{,}48;8{,}12)]$$
$$R = \frac{\sqrt{6{,}6 \ -(9{,}35)^2 \ /3}}{3}$$
$$R = 0{,}36$$

Setelah subCluster CF1, CF2, dan CF3 masuk, kemudian subCluster CF4 akan mulai masuk juga kedalam Cluster CF dan akan ada pembaharuan informasi mengenai CF. setelah bergabung dengan CF maka akan di lakukan pengecekkan terhadap radius untuk bergabung dengan CF-Leaf (Nilai R masih tidak melebihi nilai Threshold sehingga dapat bergabung dengan CFLeaf yang sama dan akan ada pembaruhan informasi mengenai CF).



Gambar 5. Leaf Split Parent Root Node

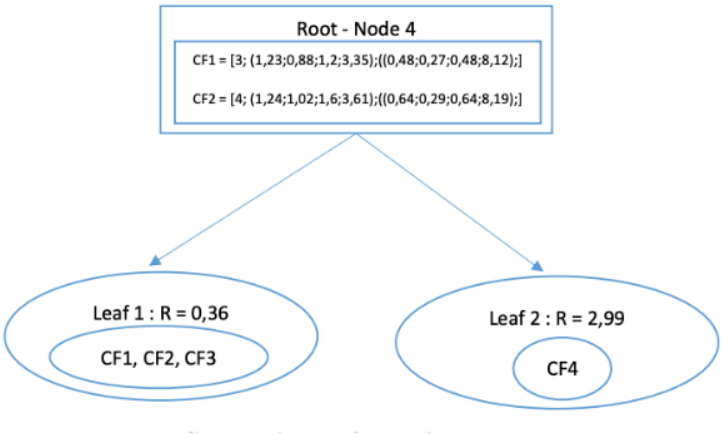
$$CF1234 = [3; (1{,}23;0{,}88;1{,}2;3{,}35);(0{,}48;0{,}27;0{,}48;8{,}12)]$$
$$R = \frac{\sqrt{7{,}87 \ -(9{,}76)^2 \ /4}}{4}$$
$$R = 2{,}99$$

Disini nilai R sudah melebihi nilai T sehingga perlu ada split pada CF Node parent. Informasi CF akan di perbaharui menjadi CF yang semula CF123 karena sudah menjadi satu Cluster

Maulana Agung Saputro: *Penulis Korespondensi

CHAIN: Journal of Computer Technology, Computer Engineering and Informatics
Website : https://ejournal.techcart-press.com/index.php/chain
E-ISSN 2964-2485
P-ISSN 2964-2450
Volume 4, Number 3, July 2026
DOI : **https://doi.org/10.58602/chain.v4i3.281**
Page 229-250

akan berubah menjadi CF1 dan menampung seluruh ringkasan dari subCluster. Kemudian akan terbentuk CF2 hasil split parent. Setelah subCluster CF1, CF2, CF3, CF4 masuk, kemudian subCluster CF5 akan mulai masuk juga kedalam Cluster CF dan akan ada pembaharuan informasi mengenai CF. Tetapi karena sudah mempunya 2 CF-Node maka, CF4 harus menentukan terlebih dahulu lokasi CF terdekat dengannya dengan menggunakan rumus distance D2. Setelah ini dapat bergabung dengan CF-leaf yang memenuhi nilai Threshold.



Gambar 6. Leaf Split Parent Root Node Kedua

Keterangan :
CF = *Cluster Feature*
N = Jumlah data di *Cluster*
LS = Jumlah koordinat data linear Pada N.
SS = Jumlah koordinat kuadrat dari data

Hitung jarak sub*Cluster* CF4 dengan CF-Node (CF1)
SS1 = 0,48+0,27+0,48+8,12 = 9,35
SS2 = 0,64+0,29+0,64+8,19 = 9,76

LS1 = 1,23+0,88+1,2+3,35 = 6,66
LS2 = 1,24+1,02+1,6+3,6 = 17,47

$$D2 = \frac{\sqrt{(N_1\, SS_2) + (N_2\, SS_1) - 2LS_1\, LS_2}}{N_1 N_2}$$

$$D2 = \frac{\sqrt{(3\,9{,}76) + (4\,9{,}76) - 2\ 6{,}66\ 7{,}47}}{3\ 4}$$

D2 = 0,88

Dari perhitungan di atas, didapatkan bahwa subCluster CF4 lebih dekat dengan CFNode (CF1) sehingga CF4 akan masuk ke CF node di CF1. Kemudian lakukan perhitungan radius.

CF1 = [3; (2,47;1,9;2,8;6,96);((1,12;0,56;1,12;16,31);]

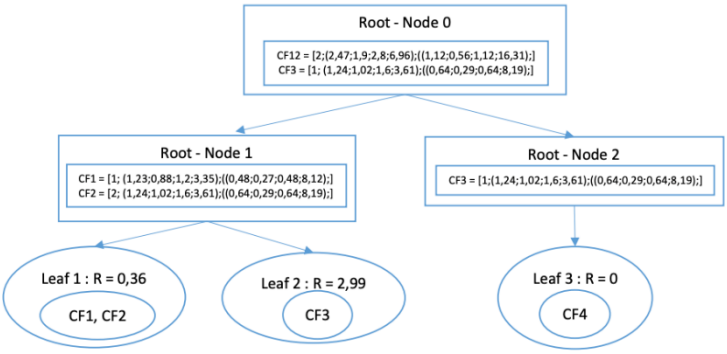
$$R = \frac{\sqrt{(14{,}13) - (19{,}11)^2/3}}{3}$$

R = 3,32

Nilai R sudah melebihi T, sehingga di perlukan leaf baru. Tetapi leaf sudah di batasi nilainya 2. (L=2) yang berarti tidak boleh ada penambahan leaf ketiga. Sehingga perlu split parent root node.

Maulana Agung Saputro: *Penulis Korespondensi

CHAIN: Journal of Computer Technology, Computer Engineering and Informatics
Website : https://ejournal.techcart-press.com/index.php/chain
E-ISSN 2964-2485
P-ISSN 2964-2450
Volume 4, Number 3, July 2026
DOI : **https://doi.org/10.58602/chain.v4i3.281**
Page 229-250



Gambar 7. Leaf Split Parent Root Node

Setelah subCluster CF1, CF2 CF3 dan CF4 masuk, kemudian subCluster CF5 akan mulai masuk juga kedalam Cluster CF dan akan ada pembaharuan informasi mengenai CF. Tetapi karena sudah mempunyai 2 CF-Node maka, CF5 harus menentukan terlebih dahulu lokasi CF terdekat dengannya dengan menggunakan rumus distance D2. Setelah ini dapat bergabung dengan CF-leaf yang memenuhi nilai Threshold.

Keterangan :

CF = Cluster Feature

N = Jumlah data di Cluster

LS= Jumlah koordinat data linear Pada N.

SS = Jumlah koordinat kuadrat dari data

Hitung jarak sub*Cluster* CF5 dengan CF-Node (CF3)

SS1 = 1,12+0,56+1,12+16,31 = 19,11

SS2 = 0,64+0,29+0,64+8,19 = 9,76

LS1 = 2,47+1,9+2,8+6,96 = 14,13

LS2 = 1,24+1,02+1,6+3,61 = 7,47

$$D2 = \frac{\sqrt{(N_1\ SS_2)+(N_2\ SS_1)-2LS_1\ LS_2}}{N_1 N_2}$$

$$D2 = \frac{\sqrt{(2\ 19,11)+(1\ 9,76)-2\ \ 14,13\ 7,47}}{3\ 4}$$

D2 = 1,04

Dari perhitungan di atas, didapatkan bahwa subCluster CF5 lebih dekat dengan CFNode (CF3) sehingga CF5 akan masuk ke CF node di CF3. Kemudian lakukan perhitungan radius. Nilai R tidak melebih nilai T sehingga subCluster CF5 dapat bergabung dengan leaf di CF3.

CF3 = [3; (2,47;1,9;2,8;6,96);((1,12;0,56;1,12;16,31);]

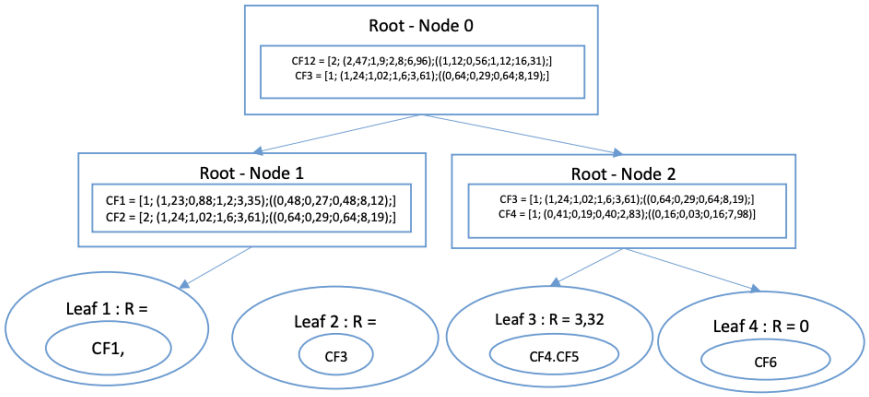
$$R = \frac{\sqrt{(14,13)-(19,11)^2/3}}{3}$$

R = 3,32

Maulana Agung Saputro: *Penulis Korespondensi

CHAIN: Journal of Computer Technology, Computer Engineering and Informatics
Website : https://ejournal.techcart-press.com/index.php/chain
E-ISSN 2964-2485
P-ISSN 2964-2450
Volume 4, Number 3, July 2026
DOI : **https://doi.org/10.58602/chain.v4i3.281**
Page 229-250

Nilai R tidak melebih nilai T sehingga subCluster CF5 dapat bergabung dengan leaf di CF3. Berikut adalah gambaran leaf tersebut :



Gambar 8. Final CF Node – Hasil Akhir Pembangunan CF-Tree

Kemudian masukkan CF7 kedalam Cluster CF dan akan ada pembaharuan informasi mengenai CF. Tetapi karena sudah mempunyai 2 CF-Node maka, CF7 harus menentukan terlebih dahulu lokasi CF terdekat dengannya dengan menggunakan rumus distance D2. Setelah ini dapat bergabung dengan CF-leaf yang memenuhi nilai Threshold.

Hitung jarak subCluster CF7 dengan CF-Node (CF12)

SS1 = 1,24+1,02+1,6+3,6 = 17,47

SS2 = 0,41+0,19+0,4+2,83 = 3,83

LS1 = 0,64+0,29+0,64+8,19 = 9,76

LS2 = 0,16+0,03+0,16+7,98 = 8,33

$$D2 = \frac{\sqrt{(N_1\, SS_2) + (N_2\, SS_1) - 2LS_1\, LS_2}}{N_1 N_2}$$

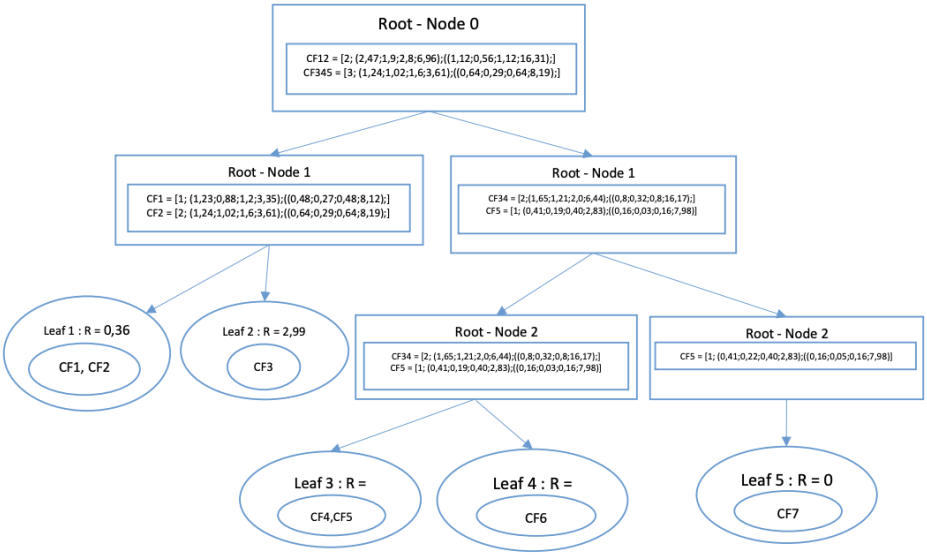
$$D2 = \frac{\sqrt{(3\,,3{,}83) + (4\,,9{,}76) - 2\ ,9{,}76\ 8{,}33}}{3\,,4}$$

$$D2 = 2.05$$

Nilai R sudah melebihi T, sehingga di perlukan leaf baru. Tetapi leaf sudah di batasi nilainya 2. (L=2) yang berarti tidak boleh ada penambahan leaf ketiga. Sehingga perlu split parent root node. Berikut dapat dilihat model CF yang terbentuk pada Gambar Final CF Node.

Maulana Agung Saputro: *Penulis Korespondensi

CHAIN: Journal of Computer Technology, Computer Engineering and Informatics
Website : https://ejournal.techcart-press.com/index.php/chain
E-ISSN 2964-2485
P-ISSN 2964-2450
Volume 4, Number 3, July 2026
DOI : **https://doi.org/10.58602/chain.v4i3.281**
Page 229-250



Gambar 9. Final CF Node

Hasil Cluster terbentuk :

CF1 : [1; (-0,41;-0,33;-0,4;-0,3);((0,16;0,11;0,16;0,07)]

R : 0,1239

CF2 : [2; (0,82;0,66;0,8;0,52);((0,32;0,22;0,32;0,14)]

R : 1,8

CF3 : [3; (1,23;0,88;1,2;3,35);((0,48;0,27;0,48;8,12)]

R : 0,36

CF4 : [3; (1,23;0,88;1,2;3,35);((0,48;0,27;0,48;8,12)]

R : 2,99

CF5 : [3; (2,47;1,9;2,8;6,96);((1,12;0,56;1,12;16,31);]

R : 3,32

CF 6 : [1; (0,41;0,22;0,40;2,83);((0,16;0,05;0,16;7,98)]

R : 0,0149

CF7 : [1; (0,41;0,37;0,40;0,26);((0,16;0,14;0,16;0,07)]

R : 0,0125

Dari hasil CF-Tree yang terbentuk, hanya ada 2 Cluster yang mendapatkan nilai subCluster lebih dekat dengan nilai R, yaitu CF4 (R = 2,99) dan CF5 (R = 3,32). Kedua cluster ini selanjutnya akan dievaluasi menggunakan metode Silhouette Coefficient.

Maulana Agung Saputro: *Penulis Korespondensi

CHAIN: Journal of Computer Technology, Computer Engineering and Informatics
Website : https://ejournal.techcart-press.com/index.php/chain
E-ISSN 2964-2485
P-ISSN 2964-2450
Volume 4, Number 3, July 2026
DOI : **https://doi.org/10.58602/chain.v4i3.281**
Page 229-250

Pengelolaan Outlier BIRCH

Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) adalah algoritma klasterisasi hierarkis yang dirancang untuk memproses data berskala besar secara efisien (Laurenso et al., 2024). Algoritma ini membangun struktur data pohon CF-Tree (Clustering Feature Tree) yang merangkum informasi statistik dari setiap sub-klaster dalam bentuk triplet CF = (N, LS, SS), di mana N adalah jumlah data poin, LS adalah jumlah linear dari data poin, dan SS adalah jumlah kuadrat dari data poin (Sari et al., 2024).

Parameter utama dalam algoritma BIRCH yang dikonfigurasi pada penelitian ini adalah:

- Threshold (T): radius maksimum sub-klaster yang menentukan apakah sebuah data poin dapat dimasukkan ke dalam sub-klaster yang ada atau membentuk sub-klaster baru. Hasan (Hasan, 2024) menekankan bahwa nilai threshold secara langsung memengaruhi distribusi dan kualitas klaster yang dihasilkan.
- Branching Factor (B): jumlah maksimum sub-klaster pada setiap node dalam CF-Tree, yang memengaruhi efisiensi memori dan waktu komputasi (Li et al., 2026).
- Jumlah Klaster (K): target jumlah klaster akhir yang dihasilkan setelah tahap klasterisasi global diterapkan pada sub-klaster dari CF-Tree (Sari et al., 2024).

Proses kerja algoritma BIRCH terdiri dari empat fase utama (Hasan, 2024): (1) Fase pemindaian data dan pembangunan CF-Tree secara inkremental; (2) Fase kondensasi CF-Tree apabila ukurannya melebihi memori yang tersedia; (3) Fase klasterisasi global menggunakan algoritma pilihan (seperti K-Means agglomerative) terhadap sub-klaster pada daun CF-Tree; dan (4) Fase penyempurnaan klaster dengan memindai ulang dataset. Pendekatan ini menjadikan BIRCH sangat cocok untuk skenario data streaming dan real-time sebagaimana ditunjukkan oleh (Li et al., 2026). Proses Pembentukan CF-Tree.

Proses clustering dimulai dengan memasukkan sub-cluster satu per satu ke dalam CF-Tree. Setiap penggabungan sub-cluster diiringi dengan pembaruan informasi CF. Ketika nilai R melampaui nilai T, akan dibentuk leaf baru. Jika jumlah leaf sudah mencapai batas maksimum L, maka akan dilakukan split pada parent root node. Proses ini berlanjut hingga seluruh data diproses dan menghasilkan Final CF Node.

Evaluasi

Validasi kualitas cluster dilakukan menggunakan Silhouette Coefficient (SC). Proses evaluasi mengukur kedekatan relasi antar objek dalam satu cluster (nilai a) dan jarak pemisahan antar cluster yang berbeda (nilai b). Nilai SC Global dihitung sebagai rata-rata dari seluruh nilai SC individual untuk mendapatkan penilaian kualitas clustering secara keseluruhan.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Evaluasi kualitas cluster pada penelitian ini dilakukan menggunakan Silhouette Coefficient (SC). Metode ini memanfaatkan dua ukuran, yaitu a_{ij} yang (merefleksikan rata-rata jarak suatu data terhadap anggota dalam cluster yang sama) dan b_{ij} yang (menunjukkan rata-

Maulana Agung Saputro: *Penulis Korespondensi

CHAIN: Journal of Computer Technology, Computer Engineering and Informatics
Website : https://ejournal.techcart-press.com/index.php/chain
E-ISSN 2964-2485
P-ISSN 2964-2450
Volume 4, Number 3, July 2026
DOI : **https://doi.org/10.58602/chain.v4i3.281**
Page 229-250

rata jarak minimum data tersebut terhadap cluster lain). Kedua komponen tersebut digunakan untuk menilai tingkat kekompakan dan keterpisahan cluster yang dihasilkan.

Deployment

Sistem deployment dibangun menggunakan bahasa pemrograman Python (versi 3) pada lingkungan Visual Studio Code. Spesifikasi hardware yang digunakan adalah Processor Intel Core i5, RAM 4 GB, dan Hardisk 1024 GB dengan sistem operasi Windows 7. Sistem menerima input file berformat CSV yang telah melalui tahap preprocessing dan menghasilkan output berupa scatter plot dan diagram visualisasi cluster.

3. HASIL DAN PEMBAHASAN

Hasil Pembentukan CF-Tree

Proses pembentukan CF-Tree dengan parameter B = 2, T = 2, dan L = 2 menghasilkan struktur Final CF Node sebagai berikut:

Tabel 11. Hasil Final CF Node

CF Node	N	CF Value (N, LS, SS)	R
CF1	1	1(-0,41,-0,33,-0,4,-0,3,0,16,0,11,0,16,0,07)	0,1239
CF2	2	2(0,82,0,66,0,8,0,52,0,32,0,22,0,32,0,14)	1,8
CF3	3	3(1,23,0,88,1,2,3,35,0,48,0,27,0,48,8,12)	0,36
CF4	3	3(1,23,0,88,1,2,3,35,0,48,0,27,0,48,8,12)	2,99
CF5	3	3(2,47,1,9,2,8,6,96,1,12,0,56,1,12,16,31)	3,32
CF6	1	1(0,41,0,22,0,40,2,83,0,16,0,05,0,16,7,98)	0,0149
CF7	1	1(0,41,0,37,0,40,0,26,0,16,0,14,0,16,0,07)	0,0125

Dari seluruh CF yang terbentuk, hanya CF4 dan CF5 yang mendapatkan nilai sub-cluster paling mendekati nilai R yang optimal. Kedua CF tersebut kemudian ditetapkan sebagai 2 cluster yang akan dievaluasi lebih lanjut menggunakan Silhouette Coefficient.

Hasil Evaluasi Silhouette Coefficient

Setelah proses pembentukan cluster menggunakan algoritma BIRCH CF-Leaf Modif, dilakukan evaluasi kualitas cluster menggunakan Silhouette Coefficient.

Tabel 12. Langkah 1 - Centroid tiap CF

Maulana Agung Saputro: *Penulis Korespondensi

CHAIN: Journal of Computer Technology, Computer Engineering and Informatics
Website : https://ejournal.techcart-press.com/index.php/chain
E-ISSN 2964-2485
P-ISSN 2964-2450
Volume 4, Number 3, July 2026
DOI : **https://doi.org/10.58602/chain.v4i3.281**
Page 229-250

CF	N	LS (d1; d2; d3; d4)	Centroid (μ1; μ2; μ3; μ4)	R (a(i))
CF1	1	−0.41 ; −0.33 ; −0.40 ; −0.30	−0.4100 ; −0.3300 ; −0.4000 ; −0.3000	0.1239
CF2	2	0.82 ; 0.66 ; 0.80 ; 0.52	0.4100 ; 0.3300 ; 0.4000 ; 0.2600	1.8
CF3	3	1.23 ; 0.88 ; 1.20 ; 3.35	0.4100 ; 0.2933 ; 0.4000 ; 1.1167	0.36
CF4	3	1.23 ; 0.88 ; 1.20 ; 3.35	0.4100 ; 0.2933 ; 0.4000 ; 1.1167	2.99
CF5	3	2.47 ; 1.90 ; 2.80 ; 6.96	0.8233 ; 0.6333 ; 0.9333 ; 2.3200	3.32
CF6	1	0.41 ; 0.22 ; 0.40 ; 2.83	0.4100 ; 0.2200 ; 0.4000 ; 2.8300	0.0149
CF7	1	0.41 ; 0.37 ; 0.40 ; 0.26	0.4100 ; 0.3700 ; 0.4000 ; 0.2600	0.0125

Tabel 13. Langkah 2 — Jarak Euclidean antar Centroid (matrix)

	CF1	CF2	CF3	CF4	CF5	CF6	CF7
CF1	0.0000	1.4358	1.9256	1.9256	3.3304	3.3781	1.4546
CF2	1.4358	0.0000	0.8575	0.8575	2.1888	2.5724	0.0400
CF3	1.9256	0.8575	0.0000	0.0000	1.4209	1.7149	0.8601
CF4	1.9256	0.8575	0.0000	0.0000	1.4209	1.7149	0.8601
CF5	3.3304	2.1888	1.4209	1.4209	0.0000	0.9414	2.1836
CF6	3.3781	2.5724	1.7149	1.7149	0.9414	0.0000	2.5744
CF7	1.4546	0.0400	0.8601	0.8601	2.1836	2.5744	0.0000

Tabel 14. Langkah 3 — a(i), b(i), dan s(i) per CF

Maulana Agung Saputro: *Penulis Korespondensi

CHAIN: Journal of Computer Technology, Computer Engineering and Informatics
Website : https://ejournal.techcart-press.com/index.php/chain
E-ISSN 2964-2485
P-ISSN 2964-2450
Volume 4, Number 3, July 2026
DOI : **https://doi.org/10.58602/chain.v4i3.281**
Page 229-250

CF	a(i) = R	b(i) = dist terdekat	CF tetangga	s(i)	Kualitas
CF1	0.1239	1.4358	CF2	0.9137	Sangat baik
CF2	1.8000	0.0400	CF7	-0.9778	Buruk
CF3	0.3600	0.0000	CF4	-1.0000	Buruk
CF4	2.9900	0.0000	CF3	-1.0000	Buruk
CF5	3.3200	0.9414	CF6	-0.7164	Buruk
CF6	0.0149	0.9414	CF5	0.9842	Sangat baik
CF7	0.0125	0.0400	CF2	0.6875	Baik

CF1 dan CF6 mendapat skor tertinggi (mendekati 1) karena radius-nya sangat kecil (0,1239 dan 0, 0149), artinya titik data di dalamnya sangat kompak, sementara jarak ke klaster lain cukup jauh.

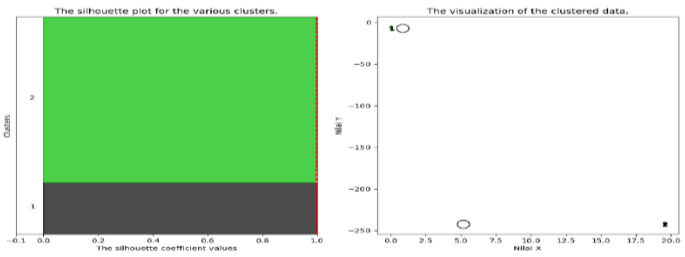
CF5 mendapat skor negatif karena radiusnya (3,32) justru lebih besar dari jarak ke centroid terdekat — artinya anggota CF5 lebih "dekat" ke klaster lain daripada ke pusatnya sendiri. Ini sinyal overlap atau threshold BIRCH yang terlalu longgar untuk area tersebut.
CF3 dan CF4 memiliki centroid yang identik (LS dan N sama) sehingga jarak antar keduanya = 0, yang menyebabkan b(i) = 0 dan s(i) = −1. Perlu diperiksa apakah CF3 dan CF4 memang merupakan dua CF yang berbeda atau ada duplikasi data input.

Karakteristik Cluster Terbentuk

Berdasarkan hasil modeling dan evaluasi, dua cluster optimal yang terbentuk memiliki karakteristik sebagai berikut:

- **Cluster 1 (Leaf 1, CF1)** — Toko dengan nilai checklist di bawah standar: mencakup toko seperti S Sukoharjo - The Park Solo - Lt. GF dengan contoh parameter checklist GLX 6128GB - Blue yang nilainya masih rendah. Cluster ini mewakili toko-toko yang memerlukan prioritas evaluasi segera.
- **Cluster 6 (Leaf 6, CF7)** — Toko dengan nilai checklist di bawah standar pada kategori berbeda: mencakup toko seperti S Sukoharjo - Hartono Mall dengan parameter checklist GLX S-Pen yang nilainya masih rendah. Cluster ini mewakili toko-toko yang memerlukan evaluasi pada aspek accessories dan product availability.

Hasil implementasi algoritma BIRCH Clustering pada data audit kepuasan pelanggan PT XYZ menghasilkan 2 cluster optimal. Proses clustering dilakukan menggunakan Python dengan parameter Branching Factor B = 2, Threshold T = 2, dan Leaf entries L = 2 pada 109.566 data audit tahun 2025.



Gambar 10. Hasil Visual BIRCH dan Silhouette Coefficient

Maulana Agung Saputro: *Penulis Korespondensi

CHAIN: Journal of Computer Technology, Computer Engineering and Informatics
Website : https://ejournal.techcart-press.com/index.php/chain
E-ISSN 2964-2485
P-ISSN 2964-2450
Volume 4, Number 3, July 2026
DOI : **https://doi.org/10.58602/chain.v4i3.281**
Page 229-250

Terlihat pada gambar 1, hasil visual BIRC dan Silhouette Coefficient sebesar 0.99930 yang mendapatkan hasil 2 Cluster yang terbentuk menggunakan algoritma BIRCH. yang dapat dilihat pada Gambar 10 Hasil Data Sampel Silhouette Coefficient. Gambar menunjukkan hasil visualisasi Silhouette Analysis dengan nilai SC = 0,99930, membuktikan bahwa 2 cluster yang terbentuk memiliki kualitas Strong Structure. Cluster 1 berisikan Toko SUKOHARJO - THE PARK SOLO - LT. GF dengan questionnaire GLX 6/128GB - Blue (26.094 data), sedangkan Cluster 2 berisikan Toko SUKOHARJO - HARTONO MALL dengan questionnaire GLX S Pen (83.474 data).

4. KESIMPULAN

Penelitian ini berhasil membangun model klasterisasi toko berdasarkan hasil monitoring summary audit kepuasan pelanggan PT XYZ menggunakan algoritma BIRCH dengan modifikasi Threshold dinamis (BIRCH CF-Leaf Modif). Dari 109.566 data audit tahun 2025 yang diproses menggunakan kerangka CRISP-DM, diperoleh 2 cluster optimal yang terbentuk dari CF4 (Leaf 4, nilai R = 2,99) dan CF5 (Leaf 5, nilai R = 3,32).

Validasi menggunakan Silhouette Coefficient menghasilkan nilai SC Cluster 1 sebesar 0,9994, SC Cluster 2 sebesar 0,9988, dan SC Global sebesar 0,9989. Nilai ini mengindikasikan bahwa cluster yang terbentuk memiliki struktur yang sangat kuat (strong structure) menurut klasifikasi Kauffman dan Rousseuw (1990), yang membuktikan efektivitas modifikasi Threshold dinamis dalam meningkatkan kualitas hasil clustering dibandingkan BIRCH dengan Threshold statis.

Secara praktis, model klasterisasi yang dihasilkan dapat digunakan SPV Toko sebagai dasar pengambilan keputusan dalam menentukan program pelatihan dan evaluasi yang tepat sasaran untuk toko-toko yang masih mendapatkan nilai final score di bawah standar perusahaan. Sistem deployment berbasis Python yang telah dibangun menghasilkan visualisasi scatter plot yang mudah dipahami oleh pengguna non-teknis.

Untuk pengembangan penelitian selanjutnya, disarankan untuk (1) melakukan modifikasi lebih lanjut pada kualitas cluster dengan eksplorasi parameter yang lebih optimal, (2) mengeksplorasi pendekatan clustering global yang berbeda untuk meningkatkan performa model, dan (3) melakukan analisis lanjutan untuk menentukan nilai Threshold yang paling ideal berdasarkan karakteristik data audit PT XYZ.

5. REFERENCES

[1] Bindini, L., Campens, L., Davis, J., Muiño-Mosquera, L., D'hulst, S., De Backer, J., Nistri, S., & Frasconi, P. (2026). Towards a more reliable assessment of aortic diameters using a Bayesian Z-score. *Scientific Reports*, *16*(1). https://doi.org/10.1038/s41598-026-46006-x

[2] Das, P., Bora, D. J., Saha, S., Lee, C. C., & Mazumdar, H. (2026). A Resilient BIRCH-Based Smart Framework for Real-Time IoT Data Clustering. *CMES - Computer Modeling in Engineering and Sciences*, *147*(1). https://doi.org/10.32604/cmes.2026.079203

[3] Fujino, I., Iphar, C., & Claramunt, C. (2026). A code distribution-based trajectory representation and its application to similarity, clustering and classification of AIS big data. *Expert Systems with Applications*, *310*(December 2025), 131216. https://doi.org/10.1016/j.eswa.2026.131216

[4] Głuszczak, B., & Powroźnik, P. (2026). *Clustering Methods in Machine Learning*. *38*(September 2025), 43–50. https://medium.com/@gunkurnia/clustering-methods-in-machine-learning-86b5639dacbd

[5] Hasan, Y. (2024). Analisis Radius Pada Algoritma Birch Berdampak Terhadap

Maulana Agung Saputro: *Penulis Korespondensi

CHAIN: Journal of Computer Technology, Computer Engineering and Informatics
Website : https://ejournal.techcart-press.com/index.php/chain
E-ISSN 2964-2485
P-ISSN 2964-2450
Volume 4, Number 3, July 2026
DOI : **https://doi.org/10.58602/chain.v4i3.281**
Page 229-250

Distribusi Dan Kualitas Cluster. *KAKIFIKOM (Kumpulan Artikel Karya Ilmiah Fakultas Ilmu Komputer)*, *06*(02), 140.

[6] Laurenso, J., Jiustian, D., Fernando, F., Suhandi, V., & Rochadiani, T. H. (2024). Implementation of K-Means, Hierarchical, and BIRCH Clustering Algorithms to Determine Marketing Targets for Vape Sales in Indonesia. *Journal of Applied Informatics and Computing*, *8*(1), 62–70. https://doi.org/10.30871/jaic.v8i1.4871

[7] Li, C., Ishak, I., Ibrahim, H., Zolkepli, M., Sidi, F., & Li, C. (2026). BIRCH-AE : A Hierarchical Ensemble Framework for Scalable E-Commerce User Segmentation with Autoencoder-Enhanced Feature Learning. *IEEE Access*, *PP*, 1. https://doi.org/10.1109/ACCESS.2026.3697984

[8] Mendes, M., & Farinha, T. (2026). MIDA—Method for Industrial Data Analysis Based on CRISP-DM. *Machine Learning and Knowledge Extraction*, *8*(4), 1–16. https://doi.org/10.3390/make8040108

[9] Sari, P. D. R., Qamal, M., & Rosnita, L. (2024). Grouping Sales Levels Smartphone of Offline Store Using BIRCH Clustering Algorithm. *International Journal of Engineering, Science and Information Technology*, *4*(4), 1–12. https://doi.org/10.52088/ijesty.v4i4.558

[10] Vardakas, G., Papakostas, I., & Likas, A. (2024). *Deep Clustering Using the Soft Silhouette Score: Towards Compact and Well-Separated Clusters*. https://doi.org/10.1007/s10994-026-07026-w

[11] Zhang, Y. (2022). DBSCAN Clustering Algorithm Based on Big Data Is Applied in Network Information Security Detection. *Security and Communication Networks*, *2022*. https://doi.org/10.1155/2022/9951609

Maulana Agung Saputro: *Penulis Korespondensi