



Peningkatan Kualitas K-Means Clustering Data Audio Musik Menggunakan Transformasi TableDC

Muhammad Aksa Hermawan*, Florentina Yuni Arini

Fakultas Matematika dan Ilmu Pengetahuan Alam, Program Studi Teknik Informatika, Universitas Negeri Semarang, Semarang, Indonesia

Email: ¹*aksahermawan2913@students.unnes.ac.id, ²floyuna@mail.unnes.ac.id

Email Penulis Korespondensi: aksahermawan2913@students.unnes.ac.id

Abstrak—Klasterisasi audio music data dengan fitur tinggi biasanya mengalami penurunan kinerja akibat efek kutukan dimensi tinggi. Dataset dengan 518 fitur K-Means classical biasanya kesulitan memodelkan hubungan antar data secara nonlinear. Tujuan penelitian ini adalah untuk menganalisis implementasi teknik transformasi ruang latent TableDC dalam tahap praproses sebelum K-Means pada FMA Small dataset. Studi kasus ini berisi 8.000 lagu dengan 518 fitur audio dan dikategorikan dalam delapan genre musik. Performansi K-Means di data asli dibandingkan dengan K-Means pada space laten yang diekstrak oleh TableDC. Analisis dilakukan dengan beberapa ukuran seperti nilai Silhouette Score, Davies-Bouldin Index, Calinski-Harabasz Index, Adjusted Rand Index, inersia atau WCSS, dan jumlah iterasi. Dari hasil eksperimen dapat disimpulkan bahwa ada peningkatan persentase yang ditawarkan oleh metode tersebut. Silhouette Score meningkat sebesar 53 persen dari nilai awal 0,0249 menjadi 0,0382. Begitu pula dengan nilai ARI yang meningkat dari nilai awal 0,0876 menjadi 0,0893. Meski demikian, nilai absolut tersebut masih sangat rendah, mengindikasikan bahwa struktur kluster yang terbentuk masih lemah dan tumpang tindih secara substansial. Dalam hal ini, representasi laten berkontribusi untuk meningkatkan efisiensi konvergensi dari 63 menjadi 47 iterasi. Nilai WCSS juga semakin berkurang dari 3.433.413 menjadi 20.628. Namun, berbeda dengan kedua indikator sebelumnya, DBI dan CHI yang berbasis linear malah mendapatkan hasil yang lebih baik dibandingkan dengan model awal, yang menunjukkan masih adanya kelemahan model dalam konteks evaluasi konvensional. Secara keseluruhan transformasi TableDC terbukti mampu meningkatkan efisiensi komputasi, namun kinerjanya belum sepenuhnya memecahkan masalah pemisahan kelas yang tumpang tindih.

Kata Kunci: Autoencoder; Deep Clustering; K-Means; Pencarian Informasi Musik; TableDC

Abstract—Clustering audio music data with high features typically suffers from performance degradation due to the curse of high dimensionality. A dataset with 518 classical K-Means features typically struggles to model nonlinear relationships between data. The purpose of this study is to analyze the implementation of the TableDC latent space transformation technique in the preprocessing stage before K-Means on the FMA Small dataset. This case study contains 8,000 songs with 518 audio features and is divided into eight music genres. The performance of K-Means on the original data is compared with that of K-Means on the latent space extracted by TableDC. The analysis is performed using several metrics such as the Silhouette Score, Davies-Bouldin Index, Calinski-Harabasz Index, Adjusted Rand Index, inertia or WCSS, and the number of iterations. The experimental results indicate a percentage improvement offered by the method. The Silhouette Score increased by 53 percent from the initial value of 0.0249 to 0.0382. Similarly, the ARI value increased from the initial value of 0.0876 to 0.0893. However, these absolute values remain very low, indicating that the formed cluster structures are still weak and substantially overlapping. In this case, the latent representation contributed to increasing the convergence efficiency from 63 to 47 iterations. The WCSS value also decreased from 3,433,413 to 20,628. However, unlike the two previous indicators, the linear-based DBI and CHI actually obtained better results compared to the initial model, which demonstrates the model's weakness in the context of conventional evaluation. Overall, the TableDC transformation has been shown to improve computational efficiency, but its performance has not fully resolved the issue of overlapping class separation.

Keywords: Autoencoder; Deep Clustering; K-Means; Music Information Retrieval; TableDC

1. PENDAHULUAN

Pembangunan teknologi digital di abad ke-21 telah menghasilkan perubahan yang sangat signifikan bagi industri musik dunia. Metode konsumsi musik di masyarakat pun ikut berubah secara radikal. Jika sebelumnya pendengar memerlukan atau harus memiliki file fisik seperti kaset maupun CD, sekarang trennya sudah beralih menjadi layanan *streaming music* yang jauh lebih luas dan tanpa batas. Platform bisnis modern di era sekarang menawarkan ratusan juta judul musik, dengan ribuan lagu baru yang masih terus bertambah setiap harinya ke dalam *database* atau pangkalan data mereka. Timbunan data yang sangat besar dan masif seperti itu pastinya membutuhkan adanya penerapan teknologi kecerdasan buatan, terutama dalam proses pengelompokan data secara otomatis atau yang biasa dikenal dengan *clustering*. Pengelompokan ini sangat dibutuhkan untuk meningkatkan performa sistem rekomendasi maupun kurasi konten secara personal, serta mempermudah pembuatan *playlist* berdasarkan selera dan keinginan pengguna [1],[2],[3]. Tanpa adanya proses *clustering* yang sistematis, pencarian lagu dalam koleksi pustaka musik yang sangat besar tentu akan sangat sulit dan memakan waktu lama jika dilakukan secara manual oleh para pengguna.

Hingga saat ini, metode klasik K-Means [4] masih sangat populer dan konsisten digunakan sebagai algoritme utama dalam pendekatan *unsupervised learning*, salah satunya adalah untuk tugas klasifikasi katalog musik [5]. Hal ini disebabkan oleh beberapa keunggulannya, yaitu struktur matematisnya yang tergolong sederhana dan tingkat beban komputasi yang cukup ringan [6],[7]. Di dalam bidang penelitian *Music Information Retrieval* atau MIR, dataset *Free Music Archive* atau FMA telah lama menjadi standar *benchmark* yang esensial untuk evaluasi efektivitas algoritme pengklasteran [8]. Salah satu variasinya, yaitu subset FMA *Small*, berisi 8.000 lagu yang terdistribusi secara merata ke dalam 8 kategori genre musik yang telah dikurasi.



Tiap trek atau lagu di dalam dataset tersebut diwakili oleh 518 fitur audio tingkat rendah (*low-level features*) berwujud data tabular. Fitur ini mencakup ekstraksi metrik yang rumit seperti *Mel-frequency cepstral coefficients* (MFCC), *spectral contrast*, *chroma STFT*, dan berbagai fitur temporal lainnya. Namun, penetapan K-Means secara langsung pada ruang fitur tingkat rendah berdimensi 518 ini menimbulkan problem geometrik dan komputasi serius yang sering disebut dengan *Curse of Dimensionality* atau kutukan dimensi tinggi [9],[10]. Dengan jumlah dimensi yang begitu besar, metrik pengukuran spasial seperti jarak *Euclidean* cenderung kehilangan kemampuannya untuk membedakan objek secara tajam. Jarak antara tiap-tiap titik data cenderung menjadi seragam dan konvergen. Situasi ini diperparah oleh adanya korelasi silang yang tinggi pada beberapa bagian dari fitur audio FMA, seperti hubungan non-linier antara komponen MFCC dan energi spektral. K-Means secara umum beroperasi dengan asumsi dasar bahwa batas-batas antar kluster bersifat linier dan bulat. Sayangnya, data audio tidak memiliki pola linier tersebut; sebaliknya, fitur-fitur ini sering kali memiliki batasan yang saling tumpang tindih (*overlap*) dan penuh dengan derau (*noise*). Sehingga, sangat penting bagi kita untuk melakukan langkah reduksi dimensi melalui pemetaan ke ruang laten non-linier sebelum melakukan proses *clustering* [11],[12].

Ada sejumlah penelitian sebelumnya yang mencoba menyelesaikan problem dimensi ini, namun tinjauan kritis menunjukkan bahwa metode-metode tersebut masih menyisakan batasan spesifik ketika dihadapkan pada karakteristik data audio 518 dimensi. Cai et al [13] menggunakan teknik reduksi dimensi linier yang disebut *Principal Component Analysis* (PCA) sebelum melakukan proses *clustering*. Teknik PCA ini berbasis pada proyeksi linier ortogonal dan hanya bertujuan memaksimalkan varians global. Pada kasus fitur FMA 518 dimensi, PCA terbukti sering kali gagal secara presisi karena metode ini secara tidak sengaja membuang sinyal-sinyal akustik minor yang non-linier—yang justru sangat signifikan untuk membedakan genre—hanya karena varians sinyal tersebut dianggap kecil. Di sisi lain, Qiu et al [14] mendesain model *deep learning* yang disebut 3D-DCDAE yang berbasis pada jaringan konvolusi. Kelemahan utama model ini adalah desain arsitekturnya yang dibuat spesifik untuk menangani data visual berupa citra spektrogram spasial (2D/3D). Ketika dihadapkan pada data tabular numerik 518 dimensi dari FMA yang sudah diekstraksi, model konvolusional standar atau *autoencoder* generik tidak efektif dalam memproses hubungan antar kolom tabular tersebut. Terdapat penelitian lain juga yang dilakukan oleh Chang & Zhang [15] yang lebih fokus melakukan *clustering* berdasarkan metrik perilaku pendengar, sehingga secara sengaja mengesampingkan parameter akustik murni dari sinyal lagu itu sendiri. Penelitian selanjutnya oleh Kidanue [16] menggunakan taktik *supervised learning* untuk mengenali genre, yang membuat modelnya sangat bergantung pada ketersediaan anotasi label manual dalam jumlah besar, sehingga tidak relevan untuk masalah *unsupervised learning* murni. Adapun Ronen et al [17] merancang *framework* yang disebut DeepDPM yang dapat menghasilkan jumlah kluster dinamis. Sayangnya, arsitektur model ini membutuhkan beban komputasi eksponensial dan efektivitasnya belum pernah diuji secara spesifik terhadap karakteristik heterogen dari fitur data tabular audio.

Dalam merespons keterbatasan metode terdahulu, arsitektur baru yang diperkenalkan oleh Rauf et al [18], yaitu TableDC, menawarkan kerangka *deep clustering* yang secara khusus dirancang untuk menangani data tabular kompleks. Selama ini efektivitas TableDC memang sudah diklaim mampu menghasilkan representasi laten yang unggul [19],[20], namun kemampuannya belum pernah divalidasi pada ruang fitur audio yang memiliki dimensi tinggi. Motivasi utama penerapan TableDC dalam penelitian ini tidak hanya sekadar untuk mereduksi dimensi, tetapi sangat berkaitan erat dengan justifikasi teoretis untuk menjembatani fenomena kesenjangan semantik (*semantic gap*) [21]. Kesenjangan semantik merujuk pada perbedaan kognitif yang besar antara fitur-fitur akustik tingkat rendah dari spektrum suara dengan konsep genre musik tingkat tinggi yang diinterpretasikan oleh pendengaran manusia. Secara empiris, dua buah lagu bisa saja memiliki pola kepadatan spektrum yang nyaris sama persis, namun lagu tersebut sebenarnya berasal dari genre yang bertolak belakang secara emosional. Dalam K-Means *baseline*, hanya perhitungan jarak numerik *Euclidean* yang digunakan, sehingga algoritma ini merespons kemiripan frekuensi secara buta dan tidak bisa mempertimbangkan unsur semantik genre tersebut [22]. Oleh karena itu, pendekatan seperti K-Means tradisional yang hanya bergantung pada penghitungan jarak murni sangat tidak tepat untuk kasus ini. Sebaliknya, pendekatan kluster *deep learning* seperti TableDC [18] memiliki potensi yang jauh lebih baik. Dengan menggabungkan pendekatan jaringan *autoencoder* dan optimasi nilai kerugian secara terpadu, metode TableDC secara proaktif memaksa data untuk menciptakan representasi ruang laten yang padat, abstrak, dan lebih terstruktur [19],[20]. Model ini diharapkan mampu menarik fitur-fitur akustik yang memiliki kohesi laten agar membentuk kluster yang lebih mencerminkan batas semantik genre yang sebenarnya [23],[24]. Berdasarkan hasil tinjauan literatur secara kritis tersebut, terdapat dua celah penelitian (*research gap*) utama yang disorot dalam makalah ini. Pertama, belum ada studi komprehensif yang memvalidasi dan menganalisis kinerja penggunaan metode *deep clustering* TableDC secara spesifik pada ruang fitur tabular data audio dengan dimensi sangat tinggi (≥ 500 dimensi) seperti dataset standar MIR FMA. Kedua, tidak ditemukan adanya studi yang melakukan analisis perbandingan sistematis antara kualitas hasil klasifikasi K-Means *baseline* biasa yang beroperasi di ruang fitur asli dengan algoritma K-Means yang beroperasi pada representasi ruang laten hasil ekstraksi TableDC.

Untuk memberikan solusi atas permasalahan tersebut, terdapat tiga tujuan objektif yang ingin dicapai dalam penelitian ini. Pertama, penerapan teknik transformasi ruang laten TableDC sebagai tahap *preprocessing* yang akan digunakan sebelum proses *clustering* menggunakan algoritma K-Means pada dataset FMA *Small* yang berisi 8.000 lagu. Kedua, melakukan analisis komparasi yang mendalam mengenai kualitas struktur kluster antara metode K-Means biasa (*baseline*) dengan metode pada ruang laten yang diusulkan. Objektif ketiga adalah melakukan evaluasi kuantitatif terkait efisiensi komputasi dan kualitas kohesi kluster dengan menggunakan instrumen metrik seperti *Silhouette Score* (SS),



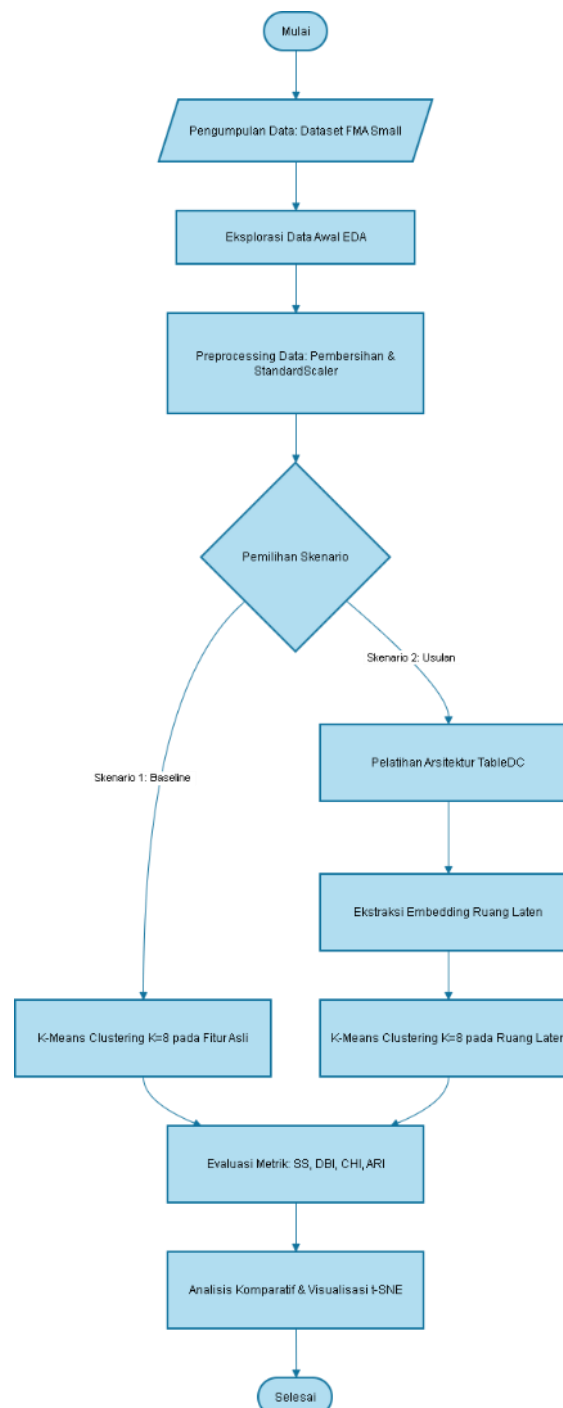
Davies-Bouldin Index (DBI), Calinski-Harabasz Index (CHI), Adjusted Rand Index (ARI), nilai inersia (WCSS), serta rasio jumlah iterasi konvergensi.

2. METODOLOGI PENELITIAN

Metodologi yang diambil dalam penelitian ini bertujuan untuk mengevaluasi kelayakan transformasi ruang laten dalam arsitektur TableDC. Fokus utama dalam metodologi ini adalah penguatan peningkatan kualitas klusterisasi K-Means berdasarkan penggunaan fitur data audio berdimensi tinggi yang memiliki karakteristik non-linier.

2.1 Tahapan Penelitian

Langkah-langkah penelitian ini akan dijalankan secara bertahap, mulai dari pengumpulan dataset, *pre-processing* fitur, reduksi data menjadi ruang laten, dan penugasan kluster. Tahapan terakhir adalah melakukan evaluasi kinerja berdasarkan berbagai metrik uji kluster. Gambar 1 memvisualisasikan seluruh proses alur kerja penelitian ini.



Gambar 1. Diagram Alir Tahapan Penelitian



Untuk menjamin keterbukaan dan keadilan proses perbandingan, kedua model penelitian dijalankan menggunakan konfigurasi inisialisasi yang identik. K-Means *clustering* yang diaplikasikan pada ruang fitur asli (*feature space*) dan K-Means *clustering* yang diaplikasikan pada representasi ruang laten TableDC sama-sama menggunakan pendekatan inisialisasi K-Means++ dengan parameter yang konsisten, yaitu $n_{init} = 10$. Pendekatan ini didasarkan pada penggunaan probabilitas penentuan pusat kluster berdasarkan jarak pusat kluster terjauh secara proporsional. Kesetaraan konfigurasi ini diimplementasikan guna menghindari adanya bias keuntungan sepihak yang tidak adil pada salah satu skenario pengujian, sehingga komparasi performa dapat dinilai secara objektif.

2.2 Sumber Data

Berkaitan dengan data, penelitian ini menggunakan dataset versi *Small* yang ada di *Free Music Archive* (FMA). Dataset tersebut sangat populer dalam bidang MIR dan telah menjadi salah satu standar uji coba [8]. Dataset ini berisi 8.000 lagu yang masing-masing memiliki panjang durasi 30 detik dan dibagi secara merata ke dalam 8 genre musik. Semua lagu telah diproses untuk mendapatkan sebanyak 518 fitur audio yang dipersembahkan dalam bentuk data tabular. Fitur ini merupakan hasil ekstraksi dari beberapa aspek seperti spektral dan dinamis ritme. Penjelasan lebih rinci tentang dataset tersebut dapat dilihat di Tabel 1.

Tabel 1. Spesifikasi Dataset FMA Varian *Small*

Parameter	Deskripsi
Jumlah Sampel	8.000 trek
Jumlah Fitur	518 dimensi
Jumlah Genre	8 (Electronic, Experimental, Folk, Hip-Hop, Instrumental, International, Pop, Rock)
Format Data	Tabular (.csv) hasil ekstraksi audio

2.3 Preprocessing Data

Tahap awal sebelum data diproses adalah penyetaraan skala menggunakan teknik *Z-score Standardization*. Langkah ini penting karena setiap fitur audio seperti *spectral centroid*, energi RMS, dan tempo memiliki satuan pengukuran dan rentang nilai yang sangat berbeda [1],[15]. Melalui proses ini, setiap fitur diubah agar memiliki nilai rata-rata 0 dan varians 1. Dengan begitu, setiap dimensi memiliki pengaruh yang seimbang saat dihitung menggunakan jarak *Euclidean* pada K-Means [16].

$$z = \frac{(x - \mu)}{\sigma} \quad (1)$$

Pada persamaan tersebut, x adalah nilai asli fitur, μ adalah nilai rata-rata dari populasi fitur, dan σ adalah simpangan baku. Proses standarisasi ini dilakukan secara menyeluruh dengan memanfaatkan operasi komputasi matriks sehingga penyesuaian bobot dapat berjalan lebih efisien tanpa merusak struktur korelasi data.

2.4 Implementasi Algoritma TableDC dan K-Means

TableDC merupakan arsitektur *deep clustering* berbasis *autoencoder* yang dirancang khusus untuk data tabular dengan distribusi kompleks [18]. Model ini memiliki dua bagian utama yaitu *encoder* dan *decoder*. *Encoder* berfungsi mengubah data mentah dari 518 dimensi menjadi representasi yang lebih ringkas yaitu 32 dimensi di ruang laten. *Decoder* kemudian bertugas merekonstruksi kembali data tersebut ke bentuk semula untuk memastikan tidak ada informasi esensial yang hilang. Proses pelatihan dilakukan dalam dua tahap. Tahap pertama adalah *pre-training* yang berfokus pada meminimalkan kesalahan rekonstruksi menggunakan *Mean Squared Error*. Tahap kedua adalah *fine-tuning* yang menggabungkan *clustering loss* berbasis *Student's t-distribution*. Tahap ini bertujuan untuk memperjelas struktur kluster dengan mendekati distribusi data ke pusat klusternya masing-masing [19].

Tabel 2. Hyperparameter Pelatihan TableDC

Hyperparameter	Nilai	Dasar Penentuan
Dimensi Latent Space	32	1/16 dari dimensi input (518), empiris
Learning Rate	0,001	Nilai umum untuk Adam optimizer
Batch Size	256	Trade-off memori vs. stabilitas gradien
Epoch Pre-Training	100	Konvergensi loss rekonstruksi stabil
Epoch Fine-Training	150	Konvergensi loss gabungan
Lambda (λ)	0,1	Bobot clustering lebih kecil dari rekonstruksi
Aktivasi Encoder	ReLU	Standar untuk non-linearitas
Random Seed	42	Reproducibility
n_{init} K-Means	10	Stabilitas inisialisasi K-Means++

Terkait penentuan *hyperparameter* (seperti jumlah *epoch*, *learning rate*, dan parameter λ), pengaturannya diadaptasi secara empiris berdasarkan rujukan standar pada literatur *deep clustering* terbaru [17],[18]. Mengingat adanya limitasi komputasi, penelitian ini secara terbuka mengakui ketiadaan proses optimasi *hyperparameter* yang sistematis—seperti penggunaan metode *Grid Search* atau *Bayesian Optimization*. Oleh karena itu, pengaturan hiperparameter ini



ditetapkan sebagai batasan metodologis eksperimen. Sebagai implikasinya, keluaran model yang dihasilkan dalam pengujian ini hanya mewakili performansi dasar (*baseline potential*) dari arsitektur TableDC, bukan merupakan gambaran performansi potensial optimalnya. Setelah proses pembentukan representasi ruang laten diselesaikan, algoritma K-Means++ diterapkan untuk membentuk 8 kluster. Jumlah ini ditetapkan secara deterministik sejak awal agar sesuai dengan jumlah label genre asli dalam dataset [22].

2.5 Evaluasi Kinerja

Tahapan terakhir adalah evaluasi hasil *clustering* dengan melihat perbandingan dari *output* hasil dengan label genre asli. Uji coba ini dijalankan menggunakan dua jenis pendekatan, yakni *internal evaluation* dan *external evaluation*. *Internal evaluation* berfungsi untuk mengevaluasi sejauh mana tingkat kemiripan dari data yang ada di dalam suatu kluster (kohesi) dan juga sejauh mana jarak antar kluster itu sendiri (separasi) tanpa perlu adanya bantuan informasi label yang sebenarnya. Beberapa instrumen ukur yang digunakan dalam kategori ini adalah *Silhouette Score*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Index* [3],[14]. Di sisi lain, *external evaluation* dilakukan dengan penggunaan *Adjusted Rand Index* sebagai instrumen ukur untuk menghitung sejauh mana keselarasan antara *output* pemisahan kluster dengan label genre *ground truth* yang tersedia [24]. Salah satu instrumen ukur paling krusial yang digunakan dalam metrik internal ini adalah *Silhouette Score*. Formula komputasinya ditampilkan dalam persamaan (2).

$$s(i) = \frac{(b(i) - a(i))}{\max(a(i), b(i))} \quad (2)$$

Di dalam rumus matematis tersebut, variabel a merupakan rata-rata jarak antara suatu titik data dengan titik-titik lainnya yang masih berada di dalam satu kluster yang sama (*intra-cluster distance*). Sedangkan variabel b merupakan jarak terdekat dari titik data tersebut menuju ke kluster tetangga lainnya (*inter-cluster distance*). Rentang nilai metrik evaluasi ini bergerak dari skala -1 sampai 1. Makin mendekati angka 1, maka tingkat pemisahan kluster dinilai makin bagus dan makin tegas pula batas partisi antar kluster tersebut.

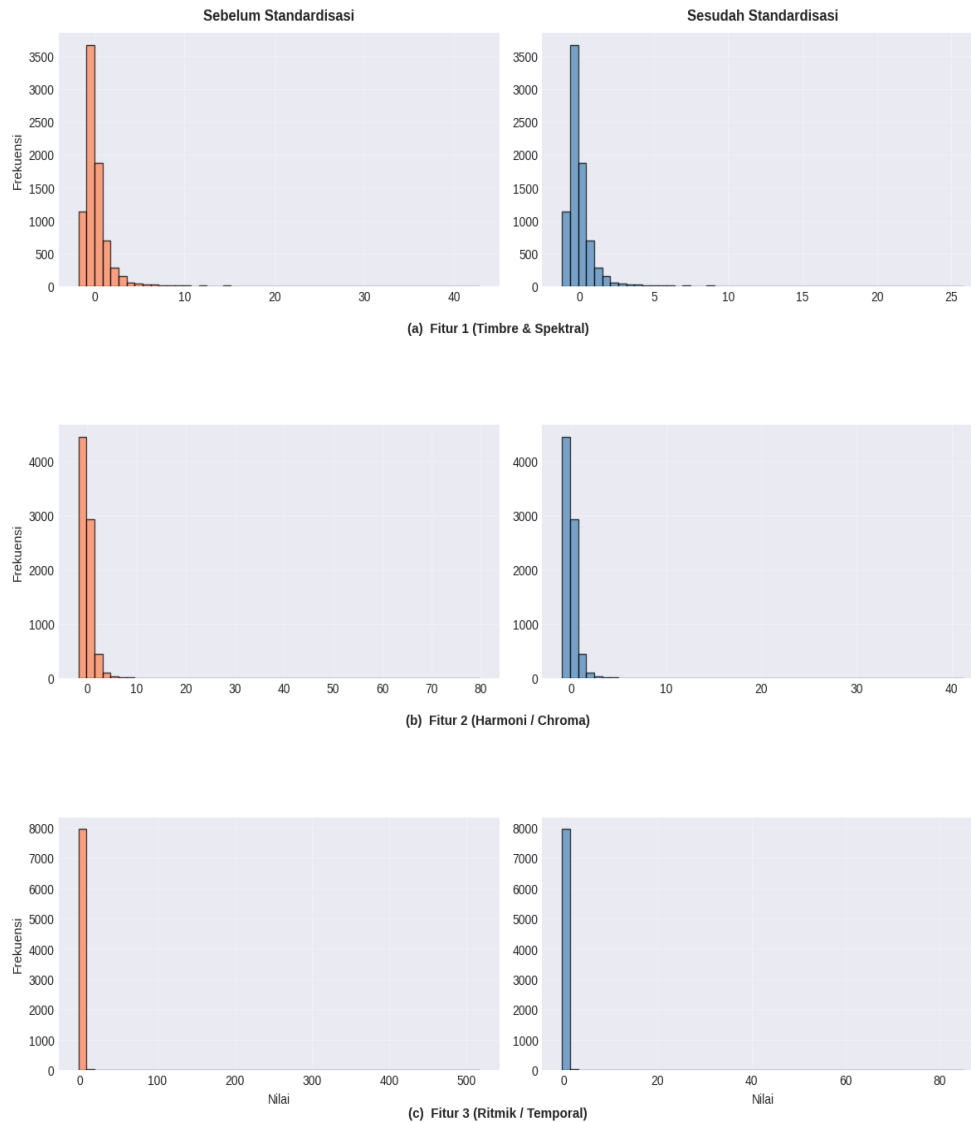
3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil evaluasi eksperimen secara menyeluruh dan terstruktur dari penerapan transformasi ruang laten menggunakan arsitektur TableDC. Transformasi ini digunakan sebagai tahap prapemrosesan sebelum algoritma K-Means dijalankan pada data audio berdimensi tinggi. Eksperimen dilakukan menggunakan dataset *Free Music Archive* varian *Small* yang berisi 8.000 trek musik dengan 8 kategori genre utama. Setiap trek dalam dataset direpresentasikan dalam 518 dimensi fitur audio. Fitur ini berasal dari ekstraksi karakteristik psikoakustik dan spektral tingkat rendah yang cukup kompleks. Di dalamnya termasuk MFCC yang menggambarkan *timbre* suara, *spectral centroid* yang menunjukkan pusat spektrum, *spectral contrast* yang menangkap perbedaan amplitudo, *chroma features* yang merepresentasikan kelas nada, serta RMSE yang mengukur kekuatan energi sinyal secara keseluruhan. Seluruh proses komputasi dalam penelitian ini bertujuan melihat sejauh mana reduksi dimensi dari 518 ke 32 dimensi melalui ruang laten non-linier dapat memperbaiki struktur data. Fokus utamanya adalah observasi perubahan pada tingkat kohesi antar data, tingkat separasi antar kluster, serta efisiensi komputasi jika dibandingkan langsung dengan eksekusi K-Means pada ruang fitur asli (*baseline*).

3.1 Hasil Preprocessing dan Eksplorasi Data

Langkah paling mendasar sebelum melakukan komputasi adalah memastikan kesiapan data serta mengetahui bagaimana pola distribusi numerik dari data tersebut. Data yang digunakan pada Dataset FMA *Small* berjumlah 518 fitur numerik yang telah diekstraksi melalui sinyal audio level rendah [8],[11]. Melihat dari analisis statistik deskriptif, diperoleh informasi bahwa setiap kelompok fitur memiliki distribusi nilai dan satuan yang sangat berbeda secara ekstrem. Ketimpangan skala ini secara langsung akan mempengaruhi performa algoritma *clustering*, baik itu K-Means konvensional maupun jaringan *Autoencoder* pada model TableDC [18]. Mengingat algoritma K-Means mengandalkan perhitungan jarak *Euclidean* absolut antar-titik, maka semakin besar rentang nilai sebuah fitur, semakin dominan pula peranannya pada hasil komputasi matriks jarak. Sebaliknya, fitur yang memiliki rentang nilai desimal kecil mungkin akan terabaikan oleh algoritma, padahal fitur tersebut bisa jadi menyimpan informasi yang sangat penting sebagai penentu genre lagu. Pada sisi model TableDC, ketidakseimbangan skala tersebut akan mempengaruhi penghitungan *loss function* secara negatif, sehingga proses pelatihan pembaruan bobot (*weight update*) menjadi tidak stabil dan tidak optimal [19]. Untuk meniyasati permasalahan ketimpangan tersebut, dilakukan proses standarisasi data secara global dengan menggunakan teknik *Z-score*. Metode matematis tersebut akan mentransformasi setiap fitur sehingga memiliki nilai rata-rata (*mean*) sebesar 0 dan deviasi standar sebesar 1 [12]. Hasil akhir dari tahapan ini adalah terbentuknya matriks data berukuran 8.000 baris dan 518 kolom yang sudah berada pada skala ekuivalen yang sama. Proses ini sangat penting karena mampu menjaga hubungan silang antar fitur tetap utuh tanpa merusak informasi aslinya, sehingga data siap digunakan untuk tahap analisis *machine learning* berikutnya.

Distribusi Fitur: Sebelum vs Sesudah Standardisasi

**Gambar 2.** Visualisasi Distribusi Fitur Audio Sebelum (Kiri) dan Sesudah (Kanan) Standardisasi Z-Score

Struktur data setelah proses standardisasi disajikan secara komparatif dalam Gambar 2. Gambar visualisasi tersebut terbagi dalam tiga bagian komparasi secara horizontal, yaitu baris (a), (b), dan (c). Untuk masing-masing bagian baris, plot distribusi di sisi kiri menunjukkan struktur data mentah asli, sedangkan plot distribusi di sisi kanan menunjukkan hasil pemerataan skala setelah proses *Z-score* diaplikasikan.

Dalam Gambar 2(a), ditampilkan perbandingan antara fitur *rmse* dan *chroma_sftf*. Pada struktur data awal di sisi kiri, fitur *rmse* memiliki batas atas nilai persebaran hingga mencapai skala 20, sedangkan batas atas dari sebaran probabilitas fitur *chroma_sftf* hanya berhenti di angka 1. Pada plot sisi kanan setelah proses standardisasi, kedua fitur tersebut kini memperoleh bentuk distribusi yang lebih seimbang dan proporsional di sekitar angka nol, menghilangkan dominasi sepihak dari fitur *rmse*.

Selanjutnya, Gambar 2(b) memvisualisasikan hasil standarisasi untuk kelompok fitur spektral, yaitu *spectral_centroid* dan *spectral_bandwidth*. Di fase data mentah awal, nilai *centroid* maksimal bisa melambung hingga melebihi angka 6.000, dan *bandwidth* bisa mencapai angka 4.000. Nilai numerik yang sangat tinggi seperti itu tentu menjadi sumber masalah utama yang bisa menimbulkan bias parah pada perhitungan jarak K-Means. Namun, melihat hasil dari proses standarisasi di sisi kanan, nilai-nilai ekstrem tersebut dapat dikontrol dan ditekan sedemikian rupa sehingga mayoritas populasinya berada pada interval stabil di sekitar rentang -2 sampai 6.

Terakhir, Gambar 2(c) memperlihatkan hasil standarisasi antara fitur *spectral_rolloff* dan *tonnetz*. Di grafik data awal, rentang nilai *rolloff* bisa membentang sangat jauh sampai menyentuh skala 10.000, sementara fitur *tonnetz* sangat terkompresi karena hanya bernilai pada pecahan desimal yang sangat kecil. Kedua fitur dengan karakter yang saling bertolak belakang tersebut menjadi jauh lebih seimbang setelah melalui proses standardisasi, dengan rentang distribusi yang berada pada area yang sama. Standardisasi *Z-score* tersebut memberikan jaminan komputasi bahwa masing-masing dari 518 fitur memiliki pijakan dan peran yang sama besarnya pada saat analisis dilakukan. Dengan distribusi yang merata,



informasi dapat diekstrak secara lebih adil dan efisien pada arsitektur *deep learning encoder* sebelum diproyeksikan pada ruang laten 32 dimensi [14]. Selain itu, K-Means konvensional juga dapat menghitung jarak *Euclidean* secara lebih objektif tanpa adanya distorsi yang didominasi oleh fitur dengan nilai metrik besar.

3.2 Hasil Evaluasi Kualitas Clustering

Pada saat matriks spasial fitur telah berada pada kondisi yang siap dan optimal untuk melakukan perhitungan, tahap clustering tanpa supervisi (*unsupervised learning*) kemudian diterapkan melalui operasi-operasi linear. Pengujian kuantitatif kemudian dilakukan melalui komparasi dua skenario kasus penelitian utama. Skenario kasus pertama merupakan penerapan algoritma K-Means yang bekerja langsung pada dataset fitur original dengan dimensi 518 (*baseline*). Skenario kasus kedua merupakan penerapan algoritma K-Means yang bekerja pada dataset fitur yang telah dipetakan ke spasial latent melalui arsitektur TableDC dengan dimensi 32 (*proposed method*). Ringkasan hasil komparasi antara kedua pendekatan ini, yang diukur menggunakan berbagai instrumen metrik standar, disajikan secara lengkap pada Tabel 3.

Tabel 3. Hasil Komparasi Metrik Performa K-Means *Baseline* dan TableDC

Metrik Evaluasi	K-Means <i>Baseline</i>	TableDC + K-Means	Δ (Perubahan)	Model Unggul
<i>Silhouette Score</i> (\uparrow)	0,0249	0,0382	+53,27%	TableDC
<i>Davies-Bouldin Index</i> (\downarrow)	3,3220	3,4918	+5,11%	K-Means
<i>Calinski-Harabasz Index</i> (\uparrow)	236,29	218,86	-7,38%	K-Means
<i>Adjusted Rand Index</i> (\uparrow)	0,0876	0,0893	+1,89%	TableDC
<i>Inertia / WCSS</i> (\downarrow)	3.433.413,09	20.628	Reduksi >99%	TableDC
<i>Jumlah Iterasi</i> (\downarrow)	63	47	-25,4%	TableDC

Dari analisis angka yang ditunjukkan pada Tabel 3, penggunaan TableDC memberikan dampak yang sangat spesifik, terutama pada pengurangan beban kalkulasi algoritma K-Means. Kompresi spasial data menjadi 32 dimensi membantu peningkatan efisiensi proses *clustering* secara drastis. Hal tersebut dapat dilihat secara objektif dari penurunan angka *Inertia* (WCSS) dan pemangkasan jumlah iterasi yang dibutuhkan oleh algoritma untuk mencapai tahap konvergensi sentroid [6],[10]. Angka WCSS mencatatkan penurunan komputasi hingga turun dari skala 3.433.413,09 pada skenario *baseline* menjadi hanya 20.628 pada skenario ruang laten. Penurunan selisih metrik yang melebihi 99% ini merupakan indikasi matematis bahwa jarak absolut antar data di dalam kluster menjadi semakin dekat secara fisik di ruang yang baru. Hal ini didukung pula dengan percepatan konvergensi jumlah iterasi yang berkurang dari 63 siklus pembaruan lokasi menjadi hanya 47 iterasi.

Kompresi data audio tingkat rendah dari 518 dimensi menjadi 32 dimensi laten membuat struktur geometri data menjadi jauh lebih padat dan fokus. Banyak kolom fitur yang masuk dalam kategori redundan atau *noise* telah disaring, sehingga tidak lagi membebani perhitungan jarak secara keseluruhan. Hasilnya, K-Means tidak perlu lagi membuang waktu untuk memproses dimensi spasial yang kosong atau berderau, dan proses pencarian titik tengah (*centroid*) pun menjadi semakin cepat diselesaikan [17]. Sebaliknya, performa K-Means yang berfungsi langsung pada data asli tampak sangat terhambat oleh fenomena *Curse of Dimensionality*. Tingginya dimensi pada data *baseline* membuat volume ruang fitur meluas secara eksponensial, sehingga jarak pembeda antara dua buah titik vektor data musik menjadi sulit untuk dikalkulasi secara presisi. Hal ini memaksa algoritma standar untuk melakukan iterasi pencarian lokasi yang semakin panjang dan berat. TableDC dalam hal ini menjalankan fungsinya sebagai filter arsitektur cerdas yang mereduksi volume ruang menjadi lebih ringkas. Data yang dihasilkan menjadi lebih ringan untuk diproses secara komputasi. Mengandalkan ruang laten tersebut, K-Means menjadi lebih efisien dalam memisahkan kluster, yang ditandai dengan penghematan siklus komputasi hingga sebesar 25,4 persen [20],[24].

3.3 Analisis Per Metrik

Untuk merumuskan pemahaman teoretis yang lebih mendalam mengenai perbedaan performa dari model yang diuji, analisis evaluasi tidak hanya dilakukan dari kesimpulan umum secara keseluruhan, namun juga dibedah secara spesifik dan berurutan untuk tiap-tiap komponen metriknya. Penjelasan analisis ini dikonstruksi berdasarkan asas-asas fundamental dari ranah *machine learning* yang kemudian dipadukan dengan kepustakaan khusus terkait dengan studi *Music Information Retrieval* atau MIR [1],[3]. Melalui perpaduan pendekatan ini, hasil numerik yang diperoleh dapat dijabarkan lebih lanjut secara objektif tanpa kehilangan landasan teoretisnya.

3.3.1 Analisis Metrik Kohesi dan Separasi: *Silhouette Score* (SS)

Metrik *Silhouette Score* (SS) diformulasikan secara khusus untuk menghitung rasio keseimbangan normatif antara tingkat kesamaan atau kedekatan data di dalam klusternya sendiri (*intra-cluster cohesion*) yang dihadapkan dengan jarak pemisah terhadap kelompok kluster terdekat lainnya (*inter-cluster separation*). Berdasarkan hasil uji kuantitatif, didapatkan fakta bahwa dengan mengadopsi teknik *deep clustering* dari TableDC, nilai SS mengalami peningkatan persentase relatif sebesar 53,27%. Nilai rasio ini meningkat dari titik basis 0,0249 pada metode linier *baseline* menjadi 0,0382 pada pemetaan ruang laten. Perbaikan persentase ini memverifikasi bahwa penerapan jaringan *Autoencoder* TableDC telah mampu merapatkan struktur data menjadi sedikit lebih dekat menuju ke arah pusat klusternya masing-masing.



Peningkatan ini salah satunya merupakan efek fungsional dari optimasi kerugian algoritma yang memanfaatkan basis perhitungan probabilitas *Student's t-distribution* [19].

Meskipun demikian, interpretasi evaluasi mutlak menuntut tinjauan yang proporsional dan sangat objektif. Apabila dipandang dari segi nilai absolut murni, skor SS sebesar 0,0382 secara faktual merupakan angka yang sangat rendah dan berkinerja buruk karena posisinya yang mendekati ekuilibrium nol. Dalam konteks evaluasi matematis, skor di kisaran tersebut menjadi bukti valid bahwa batasan pemisah (*boundary*) antar kluster belum terbentuk dengan baik. Tumpang tindih (*overlapping*) antara satu kelompok data musik dengan kelompok lainnya masih terjadi dalam skala yang cukup masif [2]. Temuan objektif ini sesungguhnya mencerminkan masalah sekaligus tantangan epistemologis paling mendasar dalam ranah klasifikasi musik tanpa pemandu label (*unsupervised*). Sebagian besar parameter fitur akustik pembentuk struktur musik mempunyai sifat irisan yang sama di lintas genre. Tingkat tempo ketukan dan kepadatan *loudness* biasanya sangat beririsan lebar. Sebagai contoh, energi frekuensi tinggi (*High Frequency Energy*) pada musik Pop yang bernuansa cepat bisa saja sangat identik dengan struktur gelombang pola musik yang ada pada *Rock* alternatif dan juga musik *Electronic*. Mengingat tidak adanya informasi label sebagai panduan (*supervision*) untuk memandu arah pembelajaran, maka K-Means hanya bisa merespons dan bergantung secara buta kepada kemiripan fitur fisika dasar tersebut. Akibatnya, beberapa lagu dari genre yang secara komersial berbeda justru dipetakan dan ditempatkan pada titik koordinat yang nyaris tumpang tindih secara absolut dalam satu ruang kluster yang sama [16],[22].

3.3.2 Analisis Metrik Dispersi Jarak: DBI dan CHI

Beralih pada pengujian pengukuran yang berkaitan murni dengan dispersi kepadatan dan ketegasan batas geometris kluster, hasil dari eksperimen justru memberikan keunggulan yang nyata bagi *baseline* K-Means di ruang asli. Untuk pengukuran metrik *Davies-Bouldin Index* (DBI), hasil yang diperoleh oleh skenario *baseline* adalah sebesar 3,3220, sedangkan skenario ruang laten TableDC memperoleh nilai 3,4918. Mengacu pada kaidah indeks DBI, nilai yang lebih kecil mengindikasikan struktur *clustering* yang memiliki rasio pemisahan lebih baik. Oleh karena itu, hasil ini menjadikan K-Means *baseline* sebagai pendekatan yang lebih unggul. Pola serupa juga terekam kuat pada pengujian *Calinski-Harabasz Index* (CHI), di mana nilai yang diperoleh dari metode *baseline* tercatat sebesar 236,29, sementara itu TableDC hanya mampu mencapai skor 218,86. Dalam indeks CHI, nilai varians komputasi yang lebih tinggi menandakan pemisahan antar kluster yang lebih tegas, sehingga temuan ini kembali mengonfirmasi keunggulan metode *baseline* secara mutlak.

Penurunan performa angka pada pengujian DBI dan CHI ini menyoroti kelemahan model TableDC ketika dihadapkan pada instrumen evaluasi konvensional. Metrik DBI dan CHI pada dasarnya diformulasikan dengan memperhitungkan kualitas struktur data yang bersandar murni pada basis perhitungan jarak linier *Euclidean* absolut. Dalam hal ini, TableDC telah merancang ulang struktur data dengan melibatkan transformasi dimensi non-linier yang cukup kompleks di dalam jaringan pelatihannya. Data musik telah dikompresi sedemikian rupa ke dalam ruang laten 32 dimensi dengan geometri topologi yang tidak lagi bersifat linier. Namun, dalam konteks evaluasi konvensional, instrumen evaluasi linier standar ini menempatkan performa K-Means *baseline* lebih superior dalam membentuk struktur konveks bulat. Hal ini menjadi temuan objektif mengenai kelemahan evaluasi representasi ruang laten bila diuji menggunakan standar pengukuran yang secara ketat menuntut batas spasial *Euclidean* yang tegas dan linier [15],[23].

3.3.3 Analisis Akurasi Eksternal: Adjusted Rand Index (ARI)

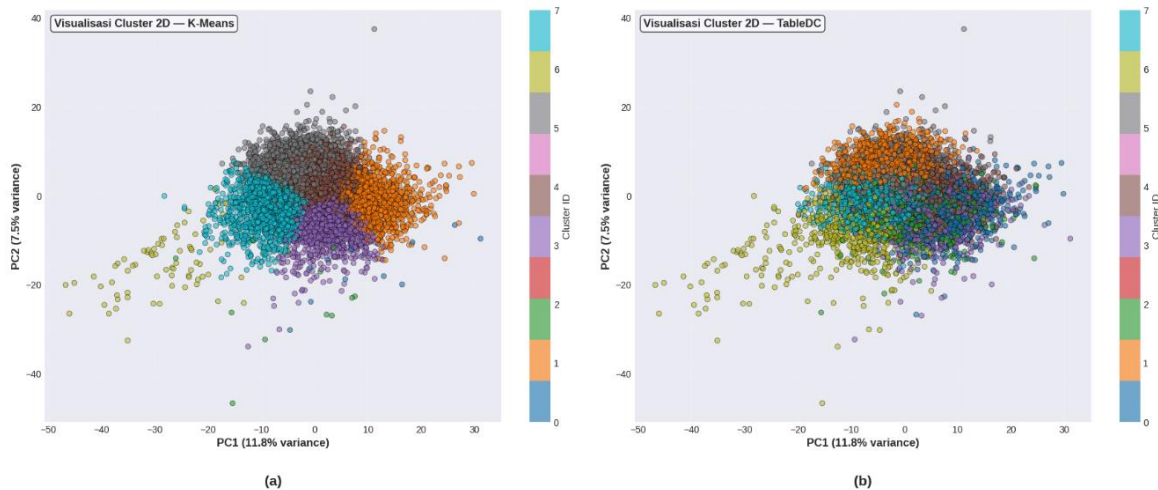
Pendekatan *external evaluation* menggunakan metrik *Adjusted Rand Index* (ARI) difungsikan sebagai metode ukur untuk mengetahui tingkat keselarasan atau korespondensi langsung antara kelompok pengklasteran algoritma yang dihasilkan dengan label *ground truth* awal. Di sinilah label yang digunakan adalah delapan jenis kategori genre murni dari dataset FMA *Small*. Metode komputasi ARI memfasilitasi penentuan kemiripan sejauh mana tebakan otonom algoritma selaras dengan klasifikasi yang dilakukan oleh kurator manusia. Adapun nilai hasil pengujian probabilitas dari ARI ini mencatatkan pergerakan naik dari 0,0876 menjadi 0,0893. Peningkatan rasio ini relatif cukup marginal, yaitu hanya sebesar 1,89 persen. Meskipun peningkatan ini menegaskan bahwa proses kompresi non-linier dari 518 fitur menjadi 32 dimensi oleh *autoencoder* tidak merusak representasi informasi utama dataset, angka absolut ARI ini tetap memerlukan telaah akademis yang kritis [7],[13].

Secara objektif, nilai indeks ARI yang berada di bawah skala 0,1 mengindikasikan bahwa hasil pemetaan K-Means bergerak pada tingkat probabilitas yang tidak jauh dari keselarasan acak (*random assignment*). Pada bidang keilmuan *Music Information Retrieval*, rendahnya batas korespondensi antara fitur mentah dan tebakan kluster genre ini sangat erat kaitannya dengan terminologi *semantic gap* (kesenjangan semantik). Istilah teoritis tersebut menjelaskan distorsi dan ketidaksetaraan yang terjadi antara pengertian sinyal energi audio secara rumusan matematis numerik dengan cara kognisi manusia dalam memaknai sebuah karya musik. Nilai fitur psikoakustik yang diekstraksi, seperti *zero crossing rate* atau rentang *mel-scale*, pada dasarnya hanyalah mencerminkan aspek fisikal bunyi secara mekanis. Di sisi lain, sebuah genre musik diklasifikasikan oleh manusia dengan melibatkan konteks budaya, kombinasi lirik tematik, sejarah musisi, serta aransemen nada instrumen tingkat tinggi. Sebagai gambaran empiris, genre musik seperti *Folk* atau *International* sering kali diklasifikasikan secara subjektif melalui unsur bahasa vokal, gaya lirik lisan, serta penggunaan alat musik tradisional lokal. Nilai-nilai subjektif semacam itu bersifat sangat abstrak, sosiologis, dan sama sekali bukan merupakan bagian yang bisa diekstrak menjadi deretan matriks data numerik. Karena itu, K-Means dan TableDC bekerja semata-mata berdasarkan angka tanpa memiliki kesadaran buatan mengenai definisi semantik sebuah genre. Keterbatasan semantik ini merupakan hambatan intrinsik yang tidak dapat dihindari sepenuhnya dalam pendekatan *unsupervised learning*. Jika ditinjau dari

realitas tersebut, angka ARI yang masih sangat rendah ini menjadi temuan valid yang menunjukkan bahwa memetakan kedekatan murni parameter suara belum mampu menyelesaikan kompleksitas dari kategorisasi *ground truth* berbasis pemahaman manusia [9],[22].

3.4 Visualisasi PCA (Principal Component Analysis)

Guna menyederhanakan pemahaman dan meningkatkan kualitas presentasi dari hasil analisis metrik kuantitatif secara visual, pengolahan matriks data yang berada pada ruang dimensi tinggi selanjutnya ditarik proyeksinya menjadi bidang datar dimensi dua (2D). Transformasi pemetaan ini bertujuan supaya pola kepadatan kluster yang ada dapat diobservasi secara langsung oleh mata manusia. Metode yang digunakan untuk memfasilitasi kebutuhan ini adalah algoritma *Principal Component Analysis* (PCA). PCA merupakan sebuah teknik reduksi dimensi populer yang beroperasi dengan menggabungkan varians maksimal data secara linier ke dalam sumbu utama ortogonal. Melalui proyeksi ini, kompleksitas dimensi dapat direduksi namun distribusi struktur data aslinya masih berupaya dipertahankan secara proporsional [13].



Gambar 3. Visualisasi Proyeksi 2D PCA dari Kluster Baseline dan TableDC

Visualisasi perbandingan proyeksi distribusi data ditunjukkan dalam dua kanvas pada Gambar 3 dengan maksud memperlihatkan tingkat perbedaan struktur kluster secara grafis. Gambar ini dibedakan menjadi dua kondisi eksibisi utama. Gambar 3(a) di sebelah kiri mengilustrasikan hasil sebaran titik dari algoritme K-Means yang mengkalkulasi ruang fitur awal. Sedangkan Gambar 3(b) menggambarkan pemetaan hasil K-Means setelah melewati ekstraksi pada ruang laten TableDC.

Dalam plot Gambar 3(a), dapat dilihat secara empiris bahwa kedelapan representasi kluster genre belum terbentuk dengan baik. Tampilan grafis ini tidak menunjukkan pemisahan gumpalan pulau data yang tegas. Tampak bahwa titik-titik data sampel musik berada dalam posisi yang saling berdekatan kuat dan ber-*overlap* satu sama lain secara acak. Bahkan, sebagian besar entitas titik terlihat berkumpul, larut, dan menggumpal pekat tepat pada bagian episentrum pusat koordinat silang. Perhatian terhadap manifestasi visual ini menegaskan realitas bahwa batas antara partisi kluster-kluster tersebut sangat cair dan kurang jelas. Fakta grafis ini secara langsung menjawab dan memvalidasi alasan di balik skor *Silhouette Score* (0,0249) yang sebelumnya bernilai sangat rendah. Terlebih lagi, gambar gumpalan padat ini juga bertindak sebagai bukti empiris yang memperkuat keberadaan fenomena *semantic gap* pada eksperimen data audio tanpa supervisi label [21]. Visualisasi tersebut merepresentasikan keadaan nyata di mana algoritme memaksakan penggabungan kelompok trek hanya dengan mempertimbangkan kemiripan nilai fitur seperti frekuensi bunyi tingkat rendah.

Beralih menelaah dari Gambar 3(b), dapat dilihat hasil proyeksi dari data komputasi yang telah melewati proses ekstraksi jaringan reduksi dimensi menggunakan arsitektur ruang laten TableDC [18]. Meskipun secara komputasi kuantitatif model *deep clustering* tersebut terbukti sukses membuat perhitungan menjadi lebih efisien dengan adanya pemangkasan drastis dari parameter *inertia*, peta dua dimensinya tetap mengindikasikan tumpang tindih. Masih terdapat persilangan titik data yang padat dan mengiris batas antar wilayah kluster. Hasil temuan proyeksi visual ini memberikan dua pandangan konfirmasi yang objektif. Pertama, temuan grafis ini sejalan dengan realitas matematis dari nilai absolut *Silhouette Score* ruang laten yang, meskipun telah meningkat menjadi 0,0382, masih berada pada level yang sangat rendah. Angka dan visual ini secara jujur mengonfirmasi bahwa struktur klusternya memang belum terpisah dengan kokoh dari ancaman *overlap*. Kedua, fenomena visual pada kanvas 2D ini juga sangat dipengaruhi oleh batasan dan kelemahan fungsional dari instrumen analitik PCA itu sendiri [13]. Sebagai mekanisme yang beroperasi murni secara transformasi ortogonal linier, PCA memiliki kerentanan untuk tidak dapat memproyeksikan pola topologi jarak spasial non-linier dari ruang laten 32 dimensi ke dalam visualisasi dua sumbu secara presisi.

3.5 Perbandingan Kontekstual dengan Penelitian Terdahulu dan Implikasi Akademis

Untuk memberikan pijakan teoretis dan memetakan posisi kontribusi dari studi eksperimen ini terhadap riset-riset sebelumnya, metodologi yang diuraikan dalam laporan ini disajikan melalui matriks komparasi pada Tabel 4. Sesuai



dengan prinsip metodologi evaluasi yang valid, penggunaan tabel matriks literatur ini secara tegas tidak difungsikan untuk membandingkan secara langsung nilai skor metrik akhir (*head-to-head*). Hal ini dikarenakan setiap kajian literatur dilaksanakan menggunakan tipe dataset, parameter uji kalibrasi, serta domain arsitektur masalah yang sama sekali berbeda. Memaksakan perbandingan akurasi angka mentah dari dataset yang berbeda (*metric mismatch comparison*) justru akan berpotensi menghasilkan klaim keunggulan (*overclaim*) yang bias dan metodologi yang tidak selaras. Oleh karena itu, Tabel 4 secara murni digunakan untuk merujuk pada pemetaan posisi navigasi kontekstual. Fokus utama komparasi difokuskan pada telaah pendekatan struktural dari kajian sebelumnya beserta batasan teknis yang mereka temui. Berangkat dari argumen tersebut, posisi kebaruan eksperimen (*novelty position*) penelitian ini dapat terlihat lebih objektif, khususnya pada tahapan implementasi arsitektur jaringan TableDC untuk tugas ekstraksi dan prapemrosesan ruang data fitur audio yang kompleks.

Tabel 4. Perbandingan Kontekstual dengan Penelitian Terdahulu

Penelitian	Objek / Dataset	Metode	Catatan Komparasi Metodologis
[13] Cai et al.	Tren data musik	PCA + K-Means	Reduksi dimensi berbasis PCA bersifat linier, berisiko menghilangkan varians fitur non-linier pada sinyal audio.
[14] Qiu et al.	Spektrogram audio (<i>Genre</i>)	3D Convolutional Autoencoder	Dirancang untuk matriks citra visual 2D/3D (spektrogram), bukan untuk data tabular numerik 1D murni.
[18] Rauf et al.	Data tabular generik (Medis, dll.)	TableDC	Arsitektur dasar teruji baik, namun belum divalidasi pada data psikoakustik dengan tumpang tindih kelas yang ekstrem.
Penelitian ini	FMA Small (Audio tabular 518 dimensi)	K-Means Baseline	SS=0,0249; ARI=0,0876 (Kondisi <i>baseline</i> dengan beban komputasi iterasi tinggi).
Penelitian ini	FMA Small (Audio tabular 518 dimensi)	TableDC + K-Means	SS=0,0382; ARI=0,0893 (Meningkatkan efisiensi iterasi dan kompresi WCSS secara substansial).

Dari analisis matriks literatur yang disajikan pada Tabel 4, terdapat penelitian sebelumnya yang dikembangkan oleh Cai et al [13] yang telah memberi perhatian penuh untuk mendukung algoritma clustering musik dengan cara mengaplikasikan teknologi transformasi data dengan menggunakan *Principal Component Analysis* (PCA). Sayangnya, ada beberapa kekurangan yang sangat nyata pada metode tersebut, dimana cara kerjanya sangat terikat pada perhitungan matematika matriks linier, sehingga tidak dapat menjaga struktur geometris varians data non-linier. Dalam hal ini, jika peneliti menggunakan metode tersebut dalam pemrosesan matriks audio yang berhubungan non-linier kompleks dalam koefisien MFCC dan fluktuasi *spectral contrast*, arsitektur linier biasanya gagal dalam menyimpan informasi akustik non-penting yang membentuk jenis musik tersebut. Arsitektur seperti TableDC telah dirancang dan dikembangkan untuk mengatasi masalah ini melalui pendekatan fungsi *deep learning* dari jaringan saraf tiruan untuk menjaga pola non-linier secara tersembunyi. Dalam poin lain dari spektrum penelitian, karya dari Qiu et al [14] telah menggunakan metodologi pemrosesan data dengan menggunakan teknologi konvolusi dari jaringan *autoencoder*. Namun demikian, fokus pengembangan arsitektur yang dirancang khusus ini dibatasi untuk proses pengolahan data spasial spasial berformat *visual image*, misalnya proyeksi spektrogram *2D pixel*. Dalam hal ini, arsitektur konvolusi malah menjadi rentan terhadap kendala komputasional apabila dipaksa mengolah format larik data numerik tabel murni seperti dataset FMA. Di sisi lain, teknologi *deep clustering* baru-baru ini diterapkan oleh Rauf et al [18] secara revolusioner untuk tabel ini, berhasil diimplementasikan dalam jenis-jenis dataset klasifikasi *generic*. Namun, efikasinya belum pernah dikaji sejauh mana validitas teknologi ini dalam kondisi dataset musik metrik audio yang memiliki distorsi silang dan *overlapping* tinggi di antara kelas-kelas tersebut.

Dengan adanya hasil uji coba yang tergolong sangat objektif tersebut, maka bisa disimpulkan bahwa implementasi integrasi transformasi TableDC masih belum berhasil dalam melakukan pekerjaannya sebagai solusi pengklasteran pembagian genre dengan sempurna. Walaupun skenario proposal ini telah memberikan pergeseran positif secara prosen peningkatan, namun pada kenyataannya telah tercatat nilai akhir absolut dari *Silhouette Score* yang tetap terhenti pada angka 0,0382 dan *Adjusted Rand Index* sebesar 0,0893. Kondisi skor yang tetap sangat mendekati angka nol tersebut memang merupakan kenyataan yang valid sebagai penanda bahwa batas pemisahan geometris antara kluster masih mengalami *overlapping* cukup signifikan. Batasan fungsional pada pemisahan kluster ini tidak dapat dipisahkan secara hierarkis dari dominasi masalah *semantic gap* [21], di mana tingkat kesamaan nilai parameter fisik akustik secara natural tidak serta merta akan sejalan dan identik dengan sentimen pemahaman apresiasi manusia terhadap batasan genre musik sebuah lagu.

Meskipun dihadapi oleh realitas metrik evaluasi *overlapping* koherensi, integrasi arsitektur TableDC memberikan nilai benefit operasional nyata dari perspektif optimasi *load* beban komputasi. Prosedur filter pemadatan matriks fitur dari *load* awal sebanyak 518 kolom berubah menjadi hanya 32 kolom ruang latent dieksekusi dengan cukup elegan, dilengkapi juga bukti penurunan metrik WCSS (Inertia) secara eksponensial dari skala besar sebesar 3.433.413 runtuh menjadi hanya



20.628. Dua hasil reduksi *load* ruang ini dikuatkan kembali dari parameter konvergen waktu, yaitu tahap pencapaian solusi dalam siklus iterasi hitung jumlah pergeseran posisi sentroid menjadi mengalami percepatan pengurangan dari *load* awal 63 loop algoritma berubah menjadi hanya membutuhkan 47 loop saja. Menilik pada fakta teknis pengurangan *load* ini, dapat dikatakan bahwa *neural network encoder* TableDC berhasil menjalankan fungsi pra-proses reduksi dimensi secara solid. Arsitektur ini berhasil mengurangi hambatan waktu komputasi dalam proses iteratif pergeseran posisi jarak rentang sentroid tanpa membuang bagian informasi laten dari datasetnya. Berdasarkan beberapa fitur kompresi unggulan, TableDC juga dapat menjadi solusi teknologi dengan kegunaan yang dapat dimanfaatkan sebagai *preprocessing gateway* dalam hal integrasi pada sistem platform *service* layanan audio streaming bisnis, di mana efisiensi *resource* waktu dalam hal pemrosesan transaksi pengolahan stream data pengguna dengan skala jumlah data yang sangat besar merupakan hal yang sangat penting bagi industri.

4. KESIMPULAN

Dari eksperimen ini, dapat disimpulkan bahwa terdapat efisiensi komputasi dalam proses konversi ruang laten menggunakan arsitektur TableDC pada tahap pra-pemrosesan algoritma K-Means untuk dataset audio FMA *Small*. Berdasarkan hasil uji coba yang melibatkan 8.000 lagu dengan 518 fitur, terjadi peningkatan nilai pada beberapa instrumen metrik. Secara persentase relatif, terjadi kenaikan nilai *Silhouette Score* (SS) sebesar 53,27% dari 0,0249 menjadi 0,0382, serta nilai *Adjusted Rand Index* (ARI) yang meningkat dari 0,0876 menjadi 0,0893. Meskipun terdapat kenaikan persentase yang cukup signifikan, nilai absolut dari skor tersebut masih sangat dekat dengan nol. Hal ini mencerminkan bahwa batasan antar kluster belum terdefinisi dengan tepat dan masih terdapat tumpang tindih (*overlap*) genre yang cukup tinggi. Selain itu, berdasarkan indeks evaluasi *Davies-Bouldin* (DBI) dan *Calinski-Harabasz* (CHI), dapat dilihat bahwa metode K-Means *baseline* berkinerja lebih baik daripada TableDC. Meskipun memiliki kelemahan pada metrik klasik tersebut, keuntungan terbesar yang didapat dari TableDC secara nyata berkaitan dengan sisi efisiensi komputasinya. Melalui reduksi yang memadatkan dimensi fitur dari 518 menjadi 32, nilai WCSS berkurang secara signifikan dari 3.433.413 menjadi 20.628. Hal ini juga membuat proses konvergensi semakin cepat dengan menurunnya jumlah iterasi dari 63 menjadi 47 kali. Perlu dipahami bahwa batasan metodologi dalam penelitian ini adalah penggunaan sampel data berskala kecil dan tidak adanya optimasi *hyperparameter*. Oleh karena itu, pada penelitian lanjutan di masa mendatang, pemanfaatan dataset yang lebih besar seperti FMA *Medium* dan penerapan optimasi *hyperparameter* secara sistematis sangat dianjurkan untuk dikembangkan.

REFERENCES

- [1] S. Mukhopadhyay, A. Kumar, D. Parashar, and M. Singh, "Enhanced Music Recommendation Systems: A Comparative Study of Content-Based Filtering and K-Means Clustering Approaches," *Revue d'Intelligence Artificielle*, vol. 38, no. 1, pp. 365–376, Feb. 2024, doi: 10.18280/ria.380138.
- [2] A. Hamid, S. Abdullah, G. Adi Pradana, and D. Gita Purnama, "K-Means Analysis for Music Segmentation in Playlist Recommendation," vol. 02, 2025, Accessed: Apr. 06, 2026. [Online]. Available: <https://doi.org/10.62201/Ocp3r511>
- [3] R. C. de Araujo, V. M. S. Santos, J. F. L. de Oliveira, and A. M. A. Maciel, "A Hybrid Music Recommendation System Based on K-Means Clustering and Multilayer Perceptron," in *International Conference on Enterprise Information Systems, ICEIS - Proceedings, Science and Technology Publications, Lda*, 2025, pp. 335–342. doi: 10.5220/0013436700003929.
- [4] F. Y. Arini *et al.*, "Optimizing K-Means Using Greylag Goose Optimization Algorithm for Household Energy Consumption Pattern Segmentation," *ILKOM Jurnal Ilmiah*, vol. 17, no. 3, pp. 302–311, 2025, doi: 10.33096/ilkom.v17i3.2851.302-311.
- [5] B. Hakim, F. Joanda Kaunang, C. Susanto, J. Salim, and R. Indradjaja, "Implementasi Machine Learning Dalam Pengelompokan Musik Menggunakan Algoritma K-Means Clustering," 2025. Accessed: Apr. 09, 2026. [Online]. Available: <http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index>
- [6] N. Aliya, A. Aini, R. Cinta Endynda, and H. Al Rosyid, "Jurnal Sains dan Teknologi Penerapan Algoritma K-Means untuk Klasterisasi Lagu Terpopuler 2025 Versi Spotify," *Jurnal Sains Dan Teknologi*, vol. 02, no. 3, p. 147, 2026, doi: 10.62379/jsit.v2i3.939.
- [7] J. Ha Rim, Y. Joo Hun, and H. Jeong Joo Hun Yoo, "Opera Clustering: K-means on librettos datasets," *Journal of Internet Computing and Services*, no. 2, pp. 45–52, 2022, doi: 10.7472/jksii.2022.23.2.45.
- [8] M. Defferrard, K. Benzi, P. Vandergheynst, and X. Bresson, "FMA: A Dataset For Music Analysis," Sep. 2017, [Online]. Available: <http://arxiv.org/abs/1612.01840>
- [9] A. Bhaskara and M. Wijewardena, "Distributed Clustering via LSH Based Data Partitioning," 2018. Accessed: Apr. 06, 2026. [Online]. Available: <https://proceedings.mlr.press/v80/bhaskara18a/bhaskara18a.pdf>
- [10] J. Hedström, S. Mair, and J. Wilzén, "Reducing Carbon Emissions in k-Means Clustering Using Representative Subsets Reducering av koldioxid-utsläpp från k-means med representativa delmängder," 2025. Accessed: Apr. 09, 2026. [Online]. Available: <https://www.diva-portal.org/smash/get/diva2:1975538/FULLTEXT01.pdf>
- [11] K. Zaman, M. Sah, C. Direkoglu, and M. Unoki, "A Survey of Audio Classification Using Deep Learning," *IEEE Access*, vol. 11, pp. 106620–106649, 2023, doi: 10.1109/ACCESS.2023.3318015.
- [12] J. Wang *et al.*, "Milvus: A Purpose-Built Vector Data Management System," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, Association for Computing Machinery, 2021, pp. 2614–2627. doi: 10.1145/3448016.3457550.
- [13] Z. Cai, L. Fu, and W. Li, "Research and analysis of music development based on k-means and PCA algorithm," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Dec. 2021. doi: 10.1088/1742-6596/2083/3/032044.



- [14] L. Qiu, S. Li, and Y. Sung, "3D-DCDAE: Unsupervised music latent representations learning method based on a deep 3d convolutional denoising autoencoder for music genre classification," *Mathematics*, vol. 9, no. 18, Sep. 2021, doi: 10.3390/math9182274.
- [15] Z. Zhang and J. Chang, "Clustering-based Categorization of Music Users through Unsupervised Learning," *Economics & Management Information*, Jun. 2018, doi: 10.58195/emi.2022.1006.
- [16] R. Kidanue, "Cornerstone: A Collection of Scholarly Cornerstone: A Collection of Scholarly and Creative Works for Minnesota and Creative Works for Minnesota Machine Learning-Driven Music Genre Recognition Machine Learning-Driven Music Genre Recognition," 2025. [Online]. Available: <https://cornerstone.lib.mnsu.edu/undergrad-theses-capstones-all>. <https://cornerstone.lib.mnsu.edu/undergrad-theses-capstones-all/8/>
- [17] M. Ronen, S. E. Finder, and O. Freifeld, "DeepDPM: Deep Clustering With an Unknown Number of Clusters," 2022. Accessed: Apr. 09, 2026. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2022/papers/Ronen_DeepDPM_Deep_Clustering_With_an_Unknown_Number_of_Clusters_CVPR_2022_paper.pdf
- [18] H. T. Rauf, A. Freitas, and N. W. Paton, "TableDC: Deep Clustering for Tabular Data," *Proceedings of the ACM on Management of Data*, vol. 3, no. 3, pp. 1–28, Jun. 2025, doi: 10.1145/3725366.
- [19] S. B. Rabbani, I. V. Medri, and M. D. Samad, "Deep clustering of tabular data by weighted Gaussian distribution learning," *Neurocomputing*, vol. 623, Mar. 2025, doi: 10.1016/j.neucom.2025.129359.
- [20] H. T. Rauf, "Deep Learning Architectures for Complex Data Fusion and Integration a Thesis Submitted to The University of Manchester for The Degree of Doctor of Philosophy in The Faculty of Science and Engineering 2026," 2026. Accessed: Apr. 06, 2026. [Online]. Available: https://pure.manchester.ac.uk/ws/portalfiles/portal/1782068998/FULL_TEXT.PDF
- [21] S. Chaudhuri and B. Raj, "Unsupervised Structure Discovery for Semantic Analysis of Audio," 2012. Accessed: Apr. 06, 2026. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/file/c6e19e830859f2cb9f7c8f8cacb8d2a6-Paper.pdf
- [22] N. Masuyama, Y. Toda, Y. Nojima, and H. Ishibuchi, "An Adaptive Resonance Theory-based Topological Clustering Algorithm with a Self-Adjusting Vigilance Parameter," Dec. 2025, [Online]. Available: <http://arxiv.org/abs/2511.17983>
- [23] N. Masuyama, T. Takebayashi, Y. Nojima, C. K. Loo, H. Ishibuchi, and S. Wermter, "A parameter-free adaptive resonance theory-based topological clustering algorithm capable of continual learning," *Neural Comput. Appl.*, vol. 38, no. 7, p. 217, Apr. 2026, doi: 10.1007/s00521-026-11907-5.
- [24] P. Li and Z. Fan, "Application of improved K-means algorithm in the cultivation of creative music talents under the needs of sustainable development and transformation," *International Journal of Web Engineering and Technology*, vol. 19, no. 1, pp. 4–19, 2024, doi: 10.1504/ijwet.2024.10063583.