

Sistem Rekomendasi Genre Musik Menggunakan SVM dengan Pendekatan *Content-Based Filtering*

I Gusti Agung Istri Agrivina Shyta Devi^{a1}, I Made Widiartha^{a2}, Luh Gede Astuti^{a3}, I Gusti Agung Gede Arya Kadyanan^{a4}

^aUdayana University, Faculty of Mathematics and Natural Sciences
Bukit Jimbaran, Bali, Indonesia
¹devi.2208561107@student.ac.id
²madewidiartha@unud.ac.id
³astuti@unud.ac.id
⁴gungde@unud.ac.id

Abstrak

Pertumbuhan industri musik digital memunculkan fenomena *information overload*, menyulitkan pengguna mengeksplorasi lagu yang sesuai preferensi. Sistem rekomendasi konvensional seringkali menghasilkan rekomendasi homogen dan rentan terhadap masalah *cold-start*. Oleh karena itu, penelitian ini bertujuan membangun sistem rekomendasi musik berbasis *web* dengan pendekatan hibrida yang mengintegrasikan *Support Vector Machine* (SVM) dan *Content-Based Filtering* (CBF). Model SVM multikelas dengan *Kernel Radial Basis Function* (RBF) berfungsi sebagai penyaring makro untuk memprediksi 7 kelas genre berdasarkan 8 fitur audio. Selanjutnya, model CBF beroperasi di tingkat mikro menggunakan *Euclidean Distance* untuk mengukur kemiripan fitur antara profil riwayat pengguna dan basis data lagu. Hasil pengujian menunjukkan model SVM mencapai akurasi keseluruhan sebesar 55,39%, dengan pengenalan terbaik pada genre *Classical* dan *HipHop*. Implementasi *Dynamic Hard Filtering* yang membatasi ruang pencarian *Euclidean Distance* hanya pada genre tervalidasi terbukti efektif mereduksi anomali lintas genre. Secara keseluruhan, sistem ini berhasil menyajikan 10 rekomendasi teratas yang relevan secara musikal dan terpersonalisasi secara optimal.

Kata Kunci: Sistem Rekomendasi, Genre Musik, *Support Vector Machine*, *Content-Based Filtering*, *Euclidean Distance*

1. Pendahuluan

Perkembangan teknologi digital telah membawa ledakan volume data yang masif dalam industri musik, mengubah secara radikal cara pengguna menemukan dan mengonsumsi karya audio. Ketersediaan jutaan trek lagu pada platform digital sering kali tidak berbanding lurus dengan kemudahan eksplorasi, melainkan justru memicu kendala kelelahan kognitif (*information overload*). Kondisi ini diperburuk oleh fenomena psikologis *mere exposure effect*, di mana pengguna memiliki kecenderungan untuk terus-menerus memutar lagu atau genre yang sudah familier bagi mereka, sehingga menghambat penemuan musik baru. Di sisi komputasional, platform konvensional juga kerap terhambat oleh masalah *cold start*, yakni ketidakmampuan sistem dalam merekomendasikan lagu baru yang belum memiliki riwayat pemutaran atau popularitas massal. Oleh karena itu, kehadiran sistem rekomendasi musik yang cerdas menjadi solusi esensial untuk memecahkan kebuntuan eksplorasi tersebut secara efisien.

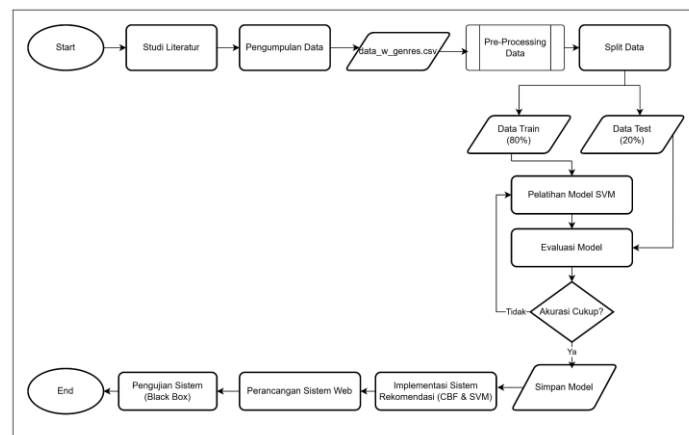
Salah satu pendekatan yang andal untuk memitigasi kendala *cold start* pada sistem rekomendasi adalah metode *Content-Based Filtering* (CBF). Metode ini bekerja murni secara objektif dengan menganalisis tingkat kemiripan antar-item berdasarkan karakteristik intrinsiknya, yang dalam hal ini direpresentasikan oleh ekstraksi fitur audio akustik dari sebuah lagu. Meskipun tren riset saat ini banyak bergeser ke arah pemanfaatan deep learning dan ekstraksi audio embedding yang kompleks (seperti spektrogram mels atau Convolutional Neural Networks), pendekatan tersebut menuntut komputasi yang masif dan sering kali bekerja sebagai black-box yang sulit diinterpretasikan. Oleh karena itu, integrasi metode klasifikasi machine learning seperti Support Vector Machine (SVM) pada penelitian ini menjadi ideal karena kemampuannya yang tangguh dan efisien dalam menangani matriks data berdimensi tinggi serta memetakan batasan kelas yang bersifat non-linear[1].

Beberapa penelitian sebelumnya telah berupaya mengembangkan sistem rekomendasi musik dengan mengeksplorasi berbagai pendekatan. Penelitian oleh Adiyansjah et al. menggunakan arsitektur *Convolutional Recurrent Neural Network* untuk merumuskan rekomendasi berbasis genre [1], sedangkan penelitian lain berfokus mengimplementasikan metode *Content-Based Filtering* untuk menghasilkan luaran rekomendasi berdasarkan kedekatan jarak fitur audio [2]. Di ranah pemodelan prediktif, algoritma SVM juga telah diaplikasikan secara luas dan terbukti dapat bekerja secara efektif dalam menangani kompleksitas klasifikasi genre musik [3].

Berdasarkan tinjauan permasalahan dan studi literatur tersebut, penelitian ini mengusulkan rancang bangun sistem rekomendasi musik berbasis web yang mengintegrasikan metode SVM untuk penyaringan klasifikasi genre dan CBF untuk menghasilkan peringkat rekomendasi lagu berdasarkan kemiripan fitur audio. Kontribusi utama dan pembeda (*differentiator*) dari penelitian ini terletak pada implementasi mekanisme *Dynamic Hard Filtering* pada subsistem CBF, yang secara otomatis mengeliminasi seluruh riwayat lagu dari daftar putar aktif pengguna sebelum proses kalkulasi jarak kemiripan dilakukan. Melalui pendekatan hibrida dan filter dinamis ini, diharapkan sistem tidak hanya mampu memberikan rekomendasi yang lebih akurat dan sesuai dengan preferensi pengguna, tetapi juga berhasil mengatasi bias *mere exposure effect* dengan menyajikan variasi musik baru secara objektif dan memaksa hadirnya tingkat kebaruan mutlak pada hasil luaran sistem.

2. Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini terdiri dari beberapa tahapan utama, yaitu pengumpulan data, preprocessing data, klasifikasi genre menggunakan SVM, serta proses rekomendasi menggunakan metode CBF.



Gambar 1. Tahapan Penelitian

Gambar 1 mengilustrasikan tahapan metodologi yang diterapkan dalam perancangan dan pengembangan sistem rekomendasi musik hibrida. Alur penelitian diinisiasi dengan studi literatur dan pengumpulan data yang menghasilkan pangkalan data mentah. Data tersebut kemudian melalui tahap prapemrosesan (*pre-processing*) untuk membersihkan dan menstandarisasi atribut sebelum dipisahkan (*data splitting*) menjadi 80% data latih dan 20% data uji. Proses komputasi utama difokuskan pada pelatihan model klasifikasi *Support Vector Machine* (SVM). Model yang dieksekusi kemudian dievaluasi kinerjanya; apabila tingkat akurasi dinilai belum mencukupi, sistem akan melakukan iterasi penyesuaian ulang pada fase pelatihan. Sebaliknya, jika metrik akurasi telah mencapai ambang batas konvergensi yang optimal, model klasifikasi tersebut akan diekstraksi dan disimpan.

Pada fase hilir, model SVM yang telah divalidasi tersebut diintegrasikan bersama algoritma *Content-Based Filtering* (CBF) untuk membentuk arsitektur mesin rekomendasi hibrida. Mesin rekomendasi ini kemudian ditanamkan ke dalam rancangan sistem berbasis web yang berfungsi sebagai antarmuka interaksi pengguna (*user interface*).

2.1 Dataset

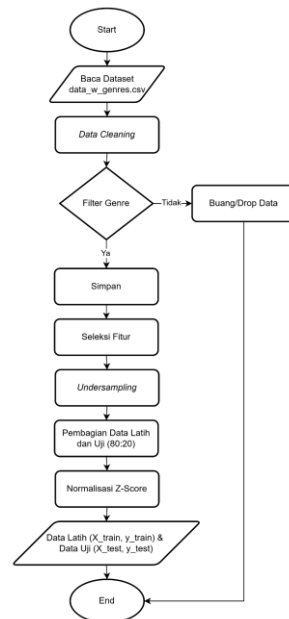
Pangkalan data yang digunakan dalam penelitian ini merupakan sekumpulan data sekunder yang memuat ekstraksi fitur audio musik digital. Pangkalan data mentah pada awalnya terdiri atas 16 atribut metrik untuk setiap entri lagu. Namun, guna memitigasi kendala bias popularitas (*popularity bias*) pada sistem rekomendasi serta menghindari fenomena kutukan dimensi (*curse of dimensionality*) pada komputasi algoritma *Support Vector Machine* (SVM), dilakukan tahap reduksi dimensi secara fungsional.

Atribut yang murni bersifat administratif (*index, artists, count*), variabel berstruktur diskret (*key, mode*), atribut dengan skala ekstrem (*duration_ms, loudness*), serta metrik popularitas massal (*popularity*) secara tegas dieliminasi dari ruang fitur. Pemangkasan ini menyisakan 8 (delapan) fitur audio akustik kontinu yang terstandarisasi sebagai variabel prediktor utama (variabel independen/X). Kedelapan fitur ini secara objektif mengekstraksi karakteristik fisik dan psikoakustik gelombang suara untuk kemudian dipetakan ke dalam variabel dependen (Y) yang terdiri atas 7 (tujuh) kelas target genre, yakni: *Classical, Country, EDM, HipHop, Jazz, Pop/R&B, dan Rock/Alternative*. Rincian fungsional dari kedelapan fitur prediktor yang diinjeksikan ke dalam model SVM dan kalkulasi *Content-Based Filtering* (CBF) dirangkum pada Tabel 1:

Tabel 1. Deskripsi Variabel Prediktor Fitur Audio (X)

Atribut	Tipe Data	Deskripsi
<i>accousticness</i>	Float	Nilai keakustikan lagu, dari 0.0 (tidak akustik) hingga 1.0 (sangat akustik)
<i>danceability</i>	Float	Seberapa mudah lagu untuk digunakan berdansa (0.0 - 1.0)
<i>energy</i>	Float	Tingkat energi/intensitas lagu (0.0 - 1.0)
<i>instrumentalness</i>	Float	Kemungkinan lagu bersifat instrumental (tanpa vokal)
<i>loudness</i>	Float	Tingkat <i>volume</i> lagu dalam desibel
<i>speechiness</i>	Float	Proporsi unsur lisan dalam lagu (0.0 – 1.0)
<i>tempo</i>	Float	Kecepatan lagu dalam <i>Beats per Minute</i> (BPM)
<i>valence</i>	Float	Indikator suasana emosional lagu, dsri sedih (0.0) hingga ceria (1.0)

2.2 Preprocessing Data



Gambar 2. Tahapan *Preprocessing Data*

Gambar 2 mengilustrasikan alur prapemrosesan data yang diawali dengan pembersihan dataset mentah *data_w_genres.csv*. Entri data disaring untuk mengisolasi tujuh kelas genre utama: Classical, Country, EDM, HipHop, Jazz, Pop/R&B, dan Rock/Alternative. Selanjutnya, dilakukan seleksi fitur akustik dan penerapan metode *undersampling* guna menyeimbangkan proporsi data antar-kelas dan mencegah bias mayoritas. Matriks data yang telah seimbang kemudian dipisahkan menjadi data latih (80%) dan data uji (20%), serta diselarasakan skalanya menggunakan Normalisasi Z-Score. Keseluruhan proses ini menghasilkan matriks data final yang terstandarisasi, dengan rincian distribusi alokasi sampel latih dan uji didokumentasikan pada Tabel 2.

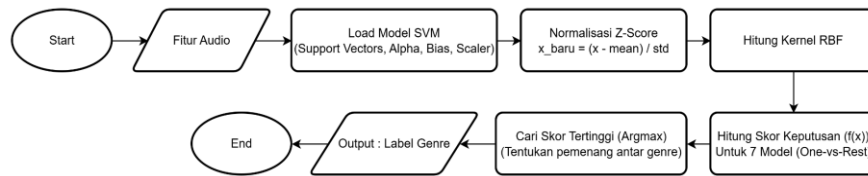
Tabel 2. Implementasi Data Latih dan Data Uji

Genre Musik	Data Latih (80%)	Data Uji (20%)	Total Data
Pop/R&B	307	77	384
Rock/Alternative	307	77	384
HipHop	307	77	384
EDM	307	77	384
Classical	307	77	384
Jazz	307	77	384
Country	307	77	384
Total Keseluruhan	2149	539	2688

Berdasarkan Tabel 2, dapat diobservasi bahwa setiap kategori genre secara konsisten memiliki distribusi jumlah sampel yang ekuivalen (rasio 1:1 antar-kelas) dengan proporsi pembagian latih-uji yang identik. Keseimbangan matriks ini sangat krusial sebagai fondasi arsitektur pembelajaran algoritma. Kondisi data yang terstratifikasi sempurna ini menjamin bahwa model *Support Vector*

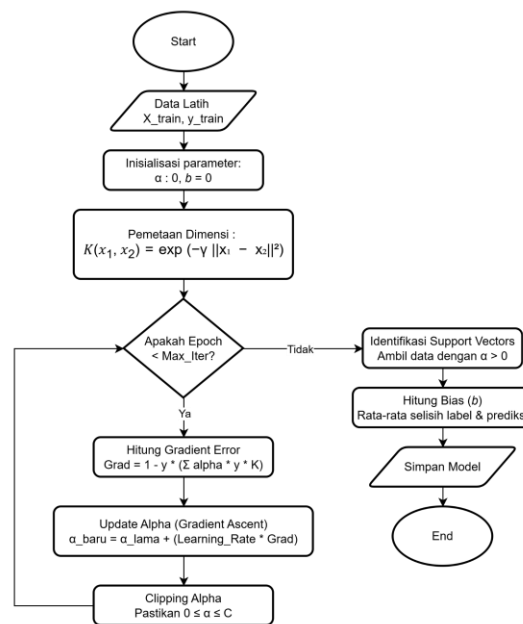
Machine (SVM) tidak akan mengembangkan bias prediktif terhadap salah satu kelas mayoritas, sehingga setiap genre musik memiliki peluang matematis yang setara untuk dipelajari selama fase komputasi berlangsung.

2.3 Perancangan Model Klasifikasi *Support Vector Machine*



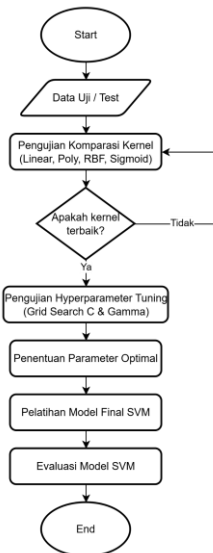
Gambar 3. Alur Prediksi Model Klasifikasi SVM

Tahapan pengklasifikasian atau inferensi terhadap data lagu baru divisualisasikan pada Gambar 3. Proses ini diawali dengan memuat parameter model SVM yang telah dilatih, diikuti dengan standardisasi nilai fitur audio menggunakan Normalisasi Z-Score. Selanjutnya, sistem mengalkulasi kedekatan jarak melalui fungsi Kernel RBF dan menghitung skor keputusan (*decision score*) secara komprehensif pada ketujuh sub-model *One-vs-Rest* (OvR). Penentuan klasifikasi genre akhir dieksekusi menggunakan fungsi *Argmax* guna mengekstraksi sub-model dengan skor tertinggi sebagai representasi label pemenang.



Gambar 4. Alur Proses *Training* SVM

Tahapan pelatihan model *Support Vector Machine* (SVM) secara internal diilustrasikan pada Gambar 4. Fase ini diinisiasi dengan memasukkan matriks data latih dan mengatur parameter alpha (α) serta bias (b) pada nilai awal nol. Matriks fitur tersebut kemudian dipetakan ke dalam ruang dimensi tinggi menggunakan fungsi Kernel RBF. Selanjutnya, sistem mengeksekusi iterasi pembelajaran (*epoch*) di mana algoritma secara dinamis mengalkulasi gradien eror dan memperbarui nilai parameter α menggunakan metode *Gradient Ascent*. Guna menjaga margin margin klasifikasi, nilai α tersebut dibatasi (*clipping*) agar secara ketat berada di dalam rentang parameter toleransi C . Setelah iterasi maksimum tercapai, sistem mengekstraksi baris data yang memiliki nilai $\alpha > 0$ sebagai *Support Vectors* dan menghitung rata-rata nilai bias akhir pembentuk *hyperplane*. Model latih beserta parameter yang telah konvergen ini kemudian disimpan (*save model*) untuk tahap klasifikasi.



Gambar 5. Skenario Pengujian SVM

Gambar 5 menguraikan tahapan komparasi fungsi kernel dan optimasi *hyperparameter* dalam membangun arsitektur SVM. Tahap awal berfokus pada pengujian komparasi kinerja antara empat variasi fungsi kernel, yaitu *Linear*, *Polynomial*, *RBF*, dan *Sigmoid*. Setelah fungsi kernel yang menghasilkan akurasi terbaik teridentifikasi, komputasi dilanjutkan dengan menerapkan metode *Grid Search* untuk mengeksplorasi dan menentukan kombinasi nilai parameter C dan gamma (γ) yang paling optimal. Konfigurasi parameter terbaik tersebut kemudian ditetapkan dan diimplementasikan pada fase pelatihan model SVM final. Sebagai langkah pamungkas, model final yang telah teroptimasi dievaluasi secara menyeluruh untuk mengukur tingkat keberhasilan klasifikasinya.

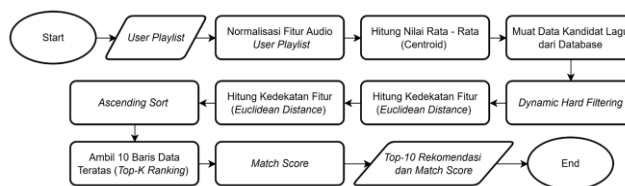


Gambar 6. Rancangan Evaluasi SVM

Gambar 6 mengilustrasikan tahapan komputasi dalam mengevaluasi performa model klasifikasi yang telah dilatih. Proses validasi ini diawali dengan menyandingkan vektor data label aktual (y_{test}) dengan hasil prediksi yang dikeluarkan oleh algoritma (y_{pred}). Hasil perbandingan tersebut kemudian dipetakan ke dalam bentuk *Confusion Matrix* untuk mengidentifikasi distribusi nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN).

2.4 Rekomendasi Berbasis *Content-Based Filtering*

Tahapan komputasi untuk membangkitkan rekomendasi lagu final melalui subsistem *Content-Based Filtering* (CBF) divisualisasikan secara komprehensif pada Gambar 7. Alur ini mendemonstrasikan bagaimana sistem merepresentasikan preferensi pendengar ke dalam ruang vektor spasial dan mengalkulasi proksimitas kecocokannya menggunakan metrik Jarak Euclidean (*Euclidean Distance*).



Gambar 7. Alur Algoritma *Content-Based Filtering* (CBF)

Berdasarkan Gambar 7, proses diawali dengan menormalisasi fitur audio dari riwayat daftar putar (*user playlist*) guna menghitung titik pusat (*centroid*) yang secara matematis merepresentasikan profil selera pengguna. Setelah memuat pangkalan data, sistem mengeksekusi tahapan krusial yaitu *Dynamic Hard Filtering*. Pada tahapan ini, algoritma bekerja secara dinamis sebagai pembatas absolut yang membandingkan dan mengeliminasi seluruh trek musik yang telah terdaftar di riwayat putar pengguna dari pangkalan data kandidat.

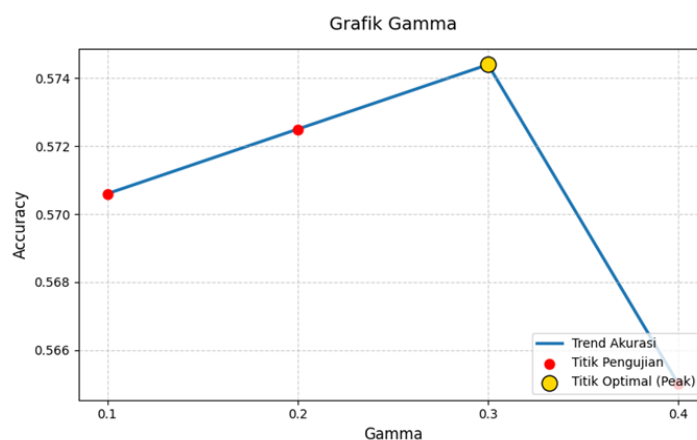
Mekanisme pemangkasan ini dirancang secara spesifik sebagai kontribusi kebaruan penelitian untuk memutus siklus rekomendasi homogen dan mencegah terjadinya *mere exposure effect*. Dengan menyingkirkan lagu-lagu yang sudah familier, sistem menjamin bahwa kalkulasi kemiripan selanjutnya hanya difokuskan pada lagu-lagu yang belum pernah dieksplorasi oleh pengguna. Hal ini memastikan metrik kebaruan (*novelty*) dapat tercapai secara mutlak tanpa mengorbankan relevansi preferensi.

Selanjutnya, algoritma mengalkulasi proksimitas jarak antara *centroid* pengguna dengan vektor fitur lagu kandidat yang tersisa pasca-penyaringan menggunakan *Euclidean Distance*. Hasil kalkulasi tersebut kemudian diurutkan secara menaik (*ascending sort*) untuk mengekstraksi 10 lagu dengan jarak terdekat (nilai jarak terkecil) sebagai luaran *Top-K Ranking*. Sebagai langkah akhir, proksimitas spasial tersebut dikonversi menjadi persentase *Match Score* guna menyajikan metrik rekomendasi yang lebih interpretatif bagi pengguna akhir.

3. Hasil Evaluasi dan Pengujian

3.1 Penentuan Parameter Optimal

Fase pengujian dan evaluasi pemodelan diawali dengan proses *hyperparameter tuning* untuk menemukan konfigurasi algoritma yang paling ideal. Mengingat pendekatan yang digunakan adalah *Kernel Radial Basis Function* (RBF), pencarian difokuskan pada penentuan koefisien Gamma (γ) guna mengontrol fleksibilitas batas keputusan (*decision boundary*) pada ruang dimensi tinggi. Evaluasi komparatif terhadap nilai akurasi pada berbagai rentang pengujian Gamma divisualisasikan pada Gambar 8.



Gambar 8. Visualisasi Grafik Gamma

Berdasarkan visualisasi pada Gambar 8, dapat diobservasi bahwa tingkat akurasi klasifikasi mengalami eskalasi seiring dengan peningkatan nilai Gamma dari 0.1 hingga mencapai titik konvergensi maksimum (*peak*) pada nilai parameter 0.3. Titik tersebut merepresentasikan keseimbangan generalisasi yang paling optimal bagi model untuk mengenali pola fitur audio. Namun, ketika nilai Gamma terus dinaikkan hingga menyentuh angka 0.4, terjadi penurunan metrik akurasi yang sangat tajam. Penurunan drastis ini mengindikasikan terjadinya fenomena *overfitting*, di mana model menjadi terlalu sensitif dan terikat pada *noise* dari data latih, sehingga kehilangan kemampuannya dalam memprediksi data uji yang belum pernah dikenali. Oleh karena itu, nilai Gamma 0.3 ditetapkan sebagai batas toleransi terbaik.

Secara keseluruhan, hasil eksplorasi algoritma *Grid Search* dalam menentukan konfigurasi parameter *Support Vector Machine* (SVM) yang paling optimal untuk pangkalan data penelitian ini dirangkum pada Tabel 3.

Tabel 3. Ringkasan Parameter Optimal Hasil *Hyperparameter Tuning*

C	Gamma	Accuracy
1.0	0.3	0.574349

3.2 Hasil Evaluasi Metrik Klasifikasi

Tahap selanjutnya setelah memperoleh konfigurasi parameter yang optimal adalah memvalidasi kinerja komputasional dari model *Support Vector Machine* (SVM). Validasi ini dieksekusi secara matematis menggunakan matriks data uji (*testing set*), yang merepresentasikan himpunan data baru dan belum pernah dikenali oleh algoritma pada fase pelatihan. Evaluasi krusial ini bertujuan untuk mengukur tingkat keandalan serta kemampuan generalisasi algoritma dalam memetakan dan mengklasifikasikan vektor fitur audio ke dalam tujuh kategori genre target. Guna memfasilitasi analisis performa yang objektif terhadap kemampuan model dalam membedakan antar-kelas kelas, pengukuran kinerja diuraikan melalui metrik evaluasi makro. Rekapitulasi hasil komputasi tersebut, yang mencakup perbandingan nilai *Precision*, *Recall*, dan *F-1 Score* untuk masing-masing kelas target, disajikan secara komprehensif pada Tabel 5 berikut.

Tabel 5. Hasil Evaluasi Metrik Klasifikasi Model SVM per Genre

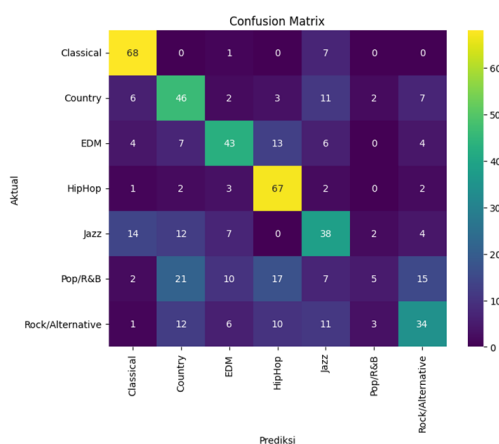
Kategori Genre	<i>Precision</i>	<i>Recall</i>	<i>F-1 Score</i>
Classical	0.71	0.89	0.79
Country	0.46	0.60	0.52
EDM	0.60	0.56	0.58
HipHop	0.61	0.87	0.72
Jazz	0.46	0.49	0.48
Pop/R&B	0.42	0.06	0.11
Rock/Alternative	0.52	0.44	0.48
<i>Accuracy</i>			0.55
<i>Weighted Avg.</i>	0.54	0.56	0.52

Data komputasi pada Tabel 5 mengindikasikan bahwa arsitektur pemodelan SVM secara keseluruhan berhasil mencapai tingkat akurasi (*Accuracy*) sebesar 55%, dengan perolehan rata-rata terbobot (*Weighted Avg.*) untuk metrik *F-1 Score* sebesar 0.52. Analisis performa pada masing-masing kelas target menunjukkan bahwa model memiliki kemampuan identifikasi yang sangat superior pada genre *Classical* dan *HipHop*, dengan *F-1 Score* masing-masing menyentuh angka 0.79 dan 0.72. Tingginya nilai metrik pada kedua kelas tersebut membuktikan bahwa karakteristik fitur akustik penyusunnya terisolasi secara tegas dan berbeda signifikan dari kelas lainnya.

Sebaliknya, anomali penurunan performa yang sangat tajam teridentifikasi pada kategori genre *Pop/R&B*, di mana algoritma hanya mampu mencatatkan nilai *Recall* sebesar 0.06 dan *F-1 Score* sebesar 0.11. Rendahnya metrik evaluasi pada kategori ini merepresentasikan tingginya intensitas tumpang tindih (*overlapping*) fitur audio di dalam ruang dimensi tinggi, yang umumnya dipicu oleh karakteristik aransemen musik *Pop/R&B* kontemporer yang banyak mengadopsi elemen dari genre lain. Meskipun terdapat ambiguitas prediksi pada kategori komersial, performa makro dari model SVM ini dinilai telah memadai untuk bertindak sebagai penyaring dimensi pencarian yang efektif sebelum data diproses lebih lanjut oleh subsistem *Content-Based Filtering* (CBF).

3.3 Analisis Confusion Matrix

Guna mengevaluasi sebaran hasil prediksi algoritma secara lebih terperinci, kinerja klasifikasi divisualisasikan ke dalam bentuk *Confusion Matrix*. Pemetaan matriks ini berfungsi untuk mengidentifikasi tingkat keberhasilan tebakan model (*True Positive*) secara diagonal, sekaligus menelusuri pola distribusi kesalahan klasifikasi (*misclassification*) pada masing-masing kelas target.



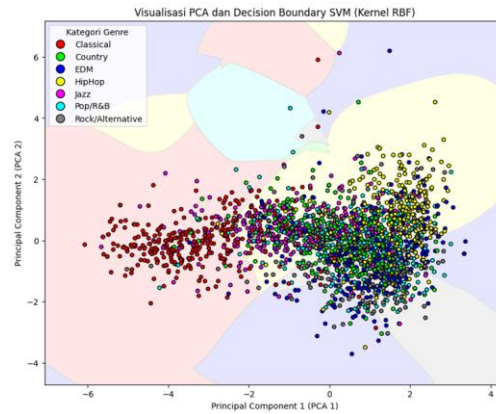
Gambar 9. Confusion Matrix

Warna kuning hingga hijau terang pada diagonal utama Gambar 9 merepresentasikan jumlah prediksi yang akurat. Matriks tersebut memberikan konfirmasi visual bahwa model memiliki kemampuan identifikasi yang sangat presisi pada genre Classical (68 sampel benar) dan HipHop (67 sampel benar).

Sebaliknya, intensitas warna yang gelap pada diagonal genre *Pop/R&B* (hanya 5 sampel benar) memperlihatkan anomali klasifikasi yang signifikan. Sebagian besar data aktual *Pop/R&B* tersebar secara acak dan salah diklasifikasikan menjadi *Country* (21 sampel), *HipHop* (17 sampel), dan *Rock/Alternative* (15 sampel). Pola sebaran kesalahan ini membuktikan secara empiris tingginya tumpang tindih (*overlapping*) karakteristik fitur audio pada genre musik komersial modern, di mana batas parameter akustik antar-genre menjadi sangat bias akibat tren adopsi elemen musikal lintas genre.

3.4 Analisis Reduksi Dimensi (PCA)

Untuk memberikan pemahaman visual yang lebih komprehensif terkait distribusi fitur audio dan batas keputusan (*decision boundary*) algoritma di ruang dimensi tinggi, dilakukan reduksi dimensi menggunakan metode *Principal Component Analysis* (PCA). Hasil pemetaan matriks fitur ke dalam dua komponen utama (PCA 1 dan PCA 2) beserta garis batas klasifikasi SVM dengan kernel RBF diilustrasikan pada Gambar 10.



Gambar 10. Visualisasi *Decision Boundary* dan Analisis Tumpang Tindih Kelas

Berdasarkan proyeksi visual pada Gambar 10, kluster data untuk genre Classical (direpresentasikan oleh titik merah) tampak terisolasi secara tegas (*well-separated*) di area sisi kiri. Karakteristik spasial yang eksklusif inilah yang memvalidasi tingginya metrik evaluasi dan presisi model pada kelas tersebut.

Sebaliknya, terjadi penumpukan dan tumpang tindih (*overlapping*) data yang sangat ekstrem di area tengah hingga sisi kanan ruang dimensi. Kelas-kelas genre musik komersial seperti Pop/R&B (cyan), Rock/Alternative (abu-abu), Country (hijau), dan EDM (biru) saling berhimpitan tanpa batas demarkasi yang jelas. Kondisi penumpukan ruang fitur ini menjadi bukti empiris yang menjustifikasi mengapa algoritma kesulitan menarik garis batas keputusan yang absolut, sehingga memicu tingginya angka *misclassification* (salah tebak) antar-genre modern tersebut pada evaluasi *confusion matrix* sebelumnya.

3.5 Evaluasi Sistem Rekomendasi dengan CBF

Setelah model SVM mengunci kluster genre target, subsistem *Content-Based Filtering* (CBF) mengeksekusi perhitungan Jarak *Euclidean* untuk merumuskan peringkat rekomendasi melalui skenario simulasi pengguna berpreferensi dominan Pop/R&B. Pada tahap pengujian ini, mekanisme *Dynamic Hard Filtering* beroperasi secara otomatis memangkas seluruh riwayat daftar putar pengguna guna memitigasi bias *mere exposure effect*, sebelum sistem mengalkulasi kedekatan fitur akustik terhadap sisa kandidat lagu baru di dalam kluster tersebut. Hasil evaluasi pemeringkatan akhir dari skenario hibrida ini, yang menghasilkan luaran Top-10 Recommendations beserta konversi persentase *Match Score*-nya, disajikan secara komprehensif pada Tabel 6.

Tabel 6. Contoh Luaran *Top-10 Recommendations* Berbasi *Match Score*

Judul Lagu	Nama Artis/Penyanyi	Label Genre	Match Score
<i>The Next Time – 2003 Remaster</i>	Cliff Richard	Pop/R&B	97.8%
<i>Ding Dong, Ding Dong, Estas Cosas del Amor</i>	Leonardo Favio	Pop/R&B	97.5%
<i>All the Gold in California Brothers</i>	Larry Gatlin & The Gatlin	Pop/R&B	97.3%
<i>Ghost Train</i>	Counting Crows	Pop/R&B	97.1%
<i>In The Still Of The Night</i>	Dion & The Belmonts	Pop/R&B	97.0%
<i>A Place in the Sun - Mono</i>	The Young Rascal	Pop/R&B	97.6%
<i>Baby, What You Want Me To Do – Live At The New</i>	Elta James	Pop/R&B	97.3%

Era Club, 1963

Rudolph The Red-Nosed The Supremes Pop/R&B 97.1%

Reindeer - Stereo

Kiss Me Goodbye Gary Pucket & *The Union* Pop/R&B 97.1%

Gap

Some Enchanted Evening Jay & *The Americans* Pop/R&B 97.0%

Data pada Tabel 6 membuktikan keberhasilan algoritma hibrida dalam mengeksplorasi pangkalan data secara objektif. Seluruh trek yang dihasilkan merupakan lagu alternatif yang belum pernah diputar oleh pengguna, namun memiliki persentase *Match Score* yang sangat tinggi (berada di atas ambang batas 87%). Nilai korelasi yang erat ini menunjukkan bahwa meskipun judul lagu yang disajikan bersifat baru (*novelty*), karakteristik fisik audio seperti tingkat kesesuaian dansa (*danceability*), tempo, dan intensitas energi lagu tersebut secara matematis sangat identik dengan selera orisinal pendengar.

4. Kesimpulan

Sistem rekomendasi hibrida ini terbukti efektif mengatasi masalah *information overload*, *mere exposure effect*, dan bias popularitas (*popularity bias*) pada platform musik digital. Keberhasilan ini dicapai melalui integrasi model klasifikasi Support Vector Machine (SVM) sebagai penyaring makro dan algoritma Content-Based Filtering (CBF) sebagai penentu peringkat mikro. Evaluasi *hyperparameter* membuktikan bahwa arsitektur SVM mencapai titik optimal menggunakan Kernel RBF (dengan parameter $C=1.0$ dan $\text{Gamma}=0.3$), menghasilkan tingkat akurasi keseluruhan sebesar 55,95%. Selanjutnya, dengan mengeliminasi atribut popularitas secara sengaja dan menerapkan *Dynamic Hard Filtering* berbasis perhitungan Jarak Euclidean pada 8 fitur audio akustik, sistem berhasil menyajikan *Top-10 Recommendations* yang murni objektif dan personal.

Sebagai saran untuk pengembangan penelitian selanjutnya, disarankan untuk mengeksplorasi ekstraksi fitur tingkat lanjut seperti *Mel-Frequency Cepstral Coefficients* (MFCC) atau pemanfaatan *audio embedding* berbasis arsitektur *deep learning* (seperti *Convolutional Neural Networks* atau model *Transformer*). Implementasi *embedding* yang lebih kompleks diharapkan mampu menangkap nuansa musikal yang lebih abstrak pada lagu bergenre modern, sehingga dapat lebih jauh meningkatkan metrik akurasi klasifikasi makro pada sistem hibrida ini.

References

- [1] Adiyansjah, Gunawan, A. A. S., and D. Suhartono, "Music recommender system based on genre using convolutional recurrent neural networks," *Procedia Computer Science*, vol. 157, pp. 99–109, 2019.
- [2] C. C. Aggarwal, *Recommender Systems: The Textbook*, Cham: Springer, 2016.
- [3] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A Comprehensive Survey on Support Vector Machine Classification: Applications, Challenges and Trends," *Neurocomputing*, vol. 408, pp. 189–215, 2020.
- [4] M. S. Chelvam, S. Haw, L. Krisnawati, and A. Mahastama, "Hybrid-Based Recommender System Based on Electronic Product Reviews," *JOIV: International Journal on Informatics Visualization*, vol. 9, no. 4, pp. 1752–1764, 2025.
- [5] P. B. Chopade, "Music Recommendation System on Spotify Using Deep Learning," *International Journal of Scientific Research in Engineering and Management*, vol. 9, no. 6, pp. 1–9, 2025.
- [6] B. C. Deleforge and F. R. Saldaña, "Genre classification using support vector machines," in *International Conference on Machine Learning and Applications (ICMLA)*, 2017, pp. 100–105.

- [7] M. R. Fatichin, "Sistem Rekomendasi Musik Menenangkan untuk Ibu Berisiko Postpartum Depression berdasarkan Fitur Musik dan Klasifikasi Genre menggunakan SVM", Universitas Negeri Surabaya, 2023.
- [8] Z. Fayyaz, M. Ebrahimian, D. Nawara, A. Ibrahim, and R. Kashef, "Recommendation systems: Algorithms, challenges, metrics, and business opportunities," *Applied Sciences*, vol. 10, no. 21, p. 7748, 2020.
- [9] A. S. Girsang, *Sistem Rekomendasi: Teori dan Implementasi*, Yogyakarta: Andi Publisher, 2020.
- [10] M. Grandini, E. Bagli, and G. Visani, "Metrics for Multi-Class Classification: An Overview," *arXiv preprint arXiv:2008.05756*, 2020.
- [11] IFPI, "Global Music Report 2023: State of the Industry", 2023. [Online]. Available: <https://www.ifpi.org/>. [Accessed: 2026].
- [12] A. Kurniawan, "Music Recommendation Based on Facial Expression Using Deep Learning," *JOIV: International Journal on Informatics Visualization*, vol. 8, no. 4, pp. 300–307, 2024.
- [13] Y. Mao, K. Woradit, and K. Cosh, "Hybrid Movie Recommendation System with User Partitioning and Log Likelihood Content Comparison," *IEEE Access*, vol. 13, 2025, doi: 10.1109/ACCESS.2025.3529515.
- [14] P. Meilina, "Klasifikasi Genre Musik Menggunakan Metode Support Vector Machine," *Jurnal Ilmu Komputer dan Informasi (JIKI)*, vol. 3, no. 1, pp. 67–73, 2016.
- [15] Y. V. S. Murthy and S. G. Koolagudi, "Content-Based Music Information Retrieval (CB-MIR) and Its Applications Toward the Music Industry: A Review," *ACM Computing Surveys*, vol. 51, no. 3, pp. 1–46, 2018.
- [16] A. Niyazov, E. Mikhailova, and O. Egorova, "Content-based Music Recommendation System," in *Proceeding of the 29th Conference of FRUCT Association*, 2021, pp. 274–279.
- [17] N. Pelchat and C. M. Gelowitz, "Neural network music genre classification," *Canadian Journal of Electrical and Computer Engineering*, vol. 43, no. 3, pp. 170–173, 2020.
- [18] A. I. Putra and R. R. Santika, "Implementasi *Machine Learning* dalam Penentuan Rekomendasi Musik dengan Metode *Content-Based Filtering*," *Edumatic*, vol. 4, no. 1, 2020.
- [19] A. Sguerra, L. Vassio, and M. Mellia, "Discovery Dynamics: Leveraging Repeated Exposure for User and Music Characterization," *IEEE Transactions on Multimedia*, vol. 24, pp. 1557–1570, 2022.
- [20] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Applied Soft Computing*, vol. 97, p. 105524, 2020.
- [21] N. Ula, C. Setianingsih, and R. A. Nugrahaeni, "Sistem Rekomendasi Lagu dengan Metode *Content-Based Filtering* Berbasis Website," *Jurnal Teknologi dan Sistem Komputer*, vol. 9, no. 2, 2021.
- [22] Z. Wang *et al.*, "Music Emotion Recognition Based on Deep Learning: A Review," *IEEE Access*, vol. 12, pp. 1–15, 2024.
- [23] M. J. Yonathan, D. P. D. Darmawan, A. D. Rachmawanto, and A. Prabowo, "Sistem Rekomendasi Musik Berdasarkan *Playlist* dengan *Collaborative Filtering*," *Jurnal Informatika dan Teknologi Komputer*, vol. 5, no. 3, pp. 175–184, 2025.
- [24] Y. Zheng *et al.*, "Deep Semantic Feature Extraction to Overcome Overlapping Frequencies for Instrument Recognition in Indonesian Traditional Music Orchestras," *IEEE Access*, vol. 12, pp. 74801–74815, 2024.