Analisis Klaster dan Klasifikasi Emosi Dalam Musik K-Pop dengan K-Means dan Algoritma C 4.5

by Reyhan Jarsi Yoga

Submission date: 12-Jul-2024 09:46AM (UTC+0700)
Submission ID: 2415503773
File name: si_Emosi_Dalam_Musik_K-Pop_dengan_K-Means_dan_Algoritma_C_4..pdf (567.81K)
Word count: 4583
Character count: 28380

Neptunus : Jurnal Ilmu Komputer Dan Teknologi Informasi
Volume. 2 No. 3 Agustus 2024
e-ISSN : 3031-898X, dan p-ISSN 3031-8998, Hal. 147-160
DOI: https://doi.org/10.61132/neptunus.v2i3.228
Available online at: https://journal.arteii.or.id/index.php/Neptunus

Analisis Klaster dan Klasifikasi Emosi Dalam Musik K-Pop dengan K-Means dan Algoritma C 4.5

[1] Reyhan Jarsi Yoga ,[2] Basuki Rahmat ,[3] Eka Prakarsa Mandyartha,
[1,2,3,] Universitas Pembangunan Nasional "Veteran" Jawa Timur

Alamat: Jl. Rungkut Madya No.1, Gn. Anyar, Kec. Gn. Anyar, Surabaya, Jawa Timur 60294
Korespondensi penulis: 20081010139@student.upnjatim.ac.id

Abstract. *The main objectives are to identify emotion patterns hidden in K-Pop music based on audio features extracted from the Spotify API and to build an emotion classification model that can predict the emotions of K-Pop songs.In this approach, the K-Means algorithm is used to cluster K-Pop songs based on audio features such as energy, valence, tempo, danceability, and speechiness. The clustering results reveal several main groups that represent variations in musical characteristics and emotions. Next, the C4.5 algorithm was used to build an emotion classification model based on the clustering results. The C4.5 model showed high performance with accuracy reaching 99.48% on a 90:10 dataset split, 99.21% on an 80:20 split, and 98.95% on a 70:30 split.The Streamlit application was developed to visualize emotion predictions from K-Pop songs with a web-based user interface. In addition, Ngrok was used to provide remote access to this application, allowing users to test and use the application remotely.The results of this study show that the combination of K-Means and C4.5 can effectively cluster and classify emotions in K-Pop music, providing valuable insights into the musical characteristics that influence emotions. This application has the potential to be used in further analysis, development of intelligent features in music applications, and improvement of user experience in listening to K-Pop music.*

Keywords: *K-Pop, cluster analysis, emotion classification, K-Means, C4.5, Spotify, Streamlit, Ngrok*

Abstrak. Tujuan utamanya adalah untuk mengidentifikasi pola emosi yang tersembunyi dalam musik K-Pop berdasarkan fitur audio yang diekstrak dari API Spotify dan untuk membangun model klasifikasi emosi yang dapat memprediksi emosi dari lagu-lagu K-Pop. Dalam pendekatan ini, algoritma K-Means digunakan untuk mengklaster lagu-lagu K-Pop berdasarkan fitur-fitur audio seperti energi, valensi, tempo, kemampuan menari, dan kemampuan berbicara. Hasil pengelompokan menunjukkan beberapa kelompok utama yang mewakili variasi karakteristik musik dan emosi. Selanjutnya, algoritma C4.5 digunakan untuk membangun model klasifikasi emosi berdasarkan hasil pengelompokan. Model C4.5 menunjukkan kinerja yang tinggi dengan akurasi mencapai 99,48% pada pembagian dataset 90:10, 99,21% pada pembagian 80:20, dan 98,95% pada pembagian 70:30. Aplikasi Streamlit dikembangkan untuk memvisualisasikan prediksi emosi dari lagu-lagu K-Pop dengan antarmuka pengguna berbasis web. Selain itu, Ngrok digunakan untuk menyediakan akses jarak jauh ke aplikasi ini, sehingga pengguna dapat menguji dan menggunakan aplikasi ini dari jarak jauh. Hasil dari penelitian ini menunjukkan bahwa kombinasi K-Means dan C4.5 dapat secara efektif mengelompokkan dan mengklasifikasikan emosi pada musik K-Pop, memberikan wawasan yang berharga tentang karakteristik musik yang mempengaruhi emosi. Aplikasi ini memiliki potensi untuk digunakan dalam analisis lebih lanjut, pengembangan fitur cerdas dalam aplikasi musik, dan peningkatan pengalaman pengguna dalam mendengarkan musik K-Pop.

Kata kunci: analisis klaster, K-Pop, klasifikasi emosi, K-Means, C4.5, Spotify, Streamlit, Ngrok,

1. LATAR BELAKANG

Musik merupakan elemen integral dalam kehidupan manusia yang berfungsi sebagai sarana ekspresi emosi, serta memiliki kemampuan untuk memengaruhi suasana hati dan kondisi mental individu. Salah satu genre musik yang mengalami perkembangan pesat dan menjadi fenomena global adalah K-Pop (Korean Pop). K-Pop memadukan berbagai elemen dari genre lain seperti pop, dance, rock, electronic pop, hip-hop, dan R&B. Fenomena ini sering disebut sebagai Korean Wave atau Hallyu, yang telah menyebar ke seluruh dunia, termasuk Indonesia (Solihah, 2016).

Received: Mei 10, 2024; Revised: Juni 15, 2024; Accepted: Juli 06, 2024; Online Available: Juli 12, 2024;
* Reyhan Jarsi Yoga , 20081010139@student.upnjatim.ac.id

Data Google Trends menunjukkan bahwa popularitas K-Pop di Indonesia meningkat secara signifikan, mendekati popularitas genre musik yang lebih mapan seperti pop, rock, jazz, dan hip-hop (Google Trends, 2024). Hal ini mencerminkan kemampuan K-Pop untuk menarik minat pendengar melalui berbagai subgenre. Contohnya, lagu "Loco" oleh Itzy memperlihatkan pengaruh pop, "Anti Fragile" oleh Lesserafim memperlihatkan genre dance, "Happy Death Day" oleh Xdinary Heroes masuk dalam genre rock, "Generation" oleh Triples mewakili electronic pop, "Baddie" oleh Ive memperkenalkan hip-hop, dan "HOLO" oleh Lee Hi menunjukkan karakteristik R&B.

Perkembangan teknologi digital telah memfasilitasi penyebaran K-Pop melalui platform streaming seperti Spotify, Apple Music, dan YouTube Music (Supriyadi, 2018). Layanan ini memungkinkan pendengar mengakses musik dari mana saja, kapan saja, dan menawarkan rekomendasi berdasarkan analisis data. Spotify, misalnya, menggunakan algoritma untuk menganalisis karakteristik audio seperti tempo, danceability, instrumentalness, loudness, speechiness, liveness, dan valence untuk menyarankan lagu-lagu serupa sesuai dengan preferensi pengguna (Ikhsan Firmansyah, Saepul Rohman, & Marsusanti, 2023). Ini menciptakan pengalaman mendengarkan yang lebih personal dan relevan.

Penelitian sebelumnya oleh Kim, Lee, dan Choi (2016) menemukan bahwa analisis klasifikasi dapat membantu mengidentifikasi hubungan antara fitur musik dan emosi yang dialami oleh pendengar. Studi mereka menunjukkan bahwa dengan menggunakan teknik klasifikasi, kita dapat lebih memahami bagaimana fitur-fitur musik tertentu mempengaruhi respons emosional. Penelitian lain oleh Nurhalimah, Hermanto, dan Kaniawulan (2022) menggunakan algoritma K-Means dan C4.5 untuk mengklasifikasikan mood musik, menyoroti bagaimana metode ini efektif dalam menentukan emosi yang terkait dengan berbagai genre musik. Kedua penelitian ini menekankan pentingnya klasifikasi dalam memahami dan mengkategorikan emosi dalam musik.

Dalam konteks K-Pop, identifikasi pola emosi berdasarkan fitur-fitur musik menjadi penting untuk meningkatkan sistem rekomendasi musik. Oleh karena itu, penelitian ini bertujuan untuk mengidentifikasi pola emosi dalam musik K-Pop dengan menggunakan algoritma klasifikasi K-Means dan algoritma pohon keputusan C4.5. K-Means adalah metode klasterisasi yang mengelompokkan data berdasarkan karakteristik tertentu, sementara algoritma C4.5 membangun pohon keputusan untuk klasifikasi data (Kaur, Dhaliwal, & Vohra, 2013). Dengan menerapkan metode ini, diharapkan dapat diperoleh pemahaman yang lebih baik mengenai bagaimana fitur analisis audio pada Spotify berkaitan dengan respons emosional pendengar (Conte, Jones, & Smith, 2020).

Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem rekomendasi musik yang lebih personal dan sesuai dengan preferensi emosional pengguna. Dengan demikian, pengalaman pendengar dalam menemukan musik yang sesuai dengan suasana hati dan preferensi mereka akan meningkat, memberikan kepuasan yang lebih besar dalam konsumsi musik.

2. KAJIAN TEORITIS

2.1. Machine Learning

Teknologi machine learning atau pembelajaran mesin adalah bidang studi dalam mengembangkan sebuah mesin untuk dapat belajar suatu hal dengan sendirinya tanpa harus diatur secara spesifik (Pratama, 2020).

Penelitian tentang pengenalan emosi dalam musik telah berkembang pesat dalam beberapa dekade terakhir. Emosi dalam musik merupakan aspek yang kompleks dan dipengaruhi oleh berbagai elemen seperti melodi, harmoni, ritme, dan lirik. Dalam konteks musik K-Pop, pengaruh budaya dan bahasa juga turut berperan dalam membentuk respon emosional pendengar (Kim et al., 2020).

2.2. Teori Pengelompokan dan Klasifikasi

Pengelompokan (clustering) adalah metode untuk mengelompokkan sekumpulan objek sedemikian rupa sehingga objek dalam kelompok yang sama (klaster) lebih mirip satu sama lain dibandingkan dengan objek di kelompok lain. Algoritma K-Means adalah salah satu metode pengelompokan yang paling populer. Dalam K-Means, jumlah klaster ditentukan terlebih dahulu, dan algoritma bekerja dengan mengiterasi centroid dari masing-masing klaster hingga tidak ada perubahan signifikan pada posisi centroid (Jain, 2010). Keuntungan utama dari K-Means adalah kesederhanaannya dan efisiensinya dalam mengelola data dalam jumlah besar (Wu et al., 2008).

Algoritma K-Means telah digunakan dalam berbagai studi untuk mengelompokkan data musik berdasarkan fitur audio seperti tempo, valensi, dan energi yang mencerminkan emosi yang berbeda (Lee et al., 2019). Dalam penelitian ini, K-Means digunakan untuk mengelompokkan musik K-Pop berdasarkan atribut audio yang diperoleh dari Spotify. Setelah pengelompokan, algoritma C4.5 diterapkan untuk klasifikasi emosi, membangun pohon keputusan yang transparan dan mudah diinterpretasikan (Quinlan, 1993).

Algoritma C4.5 adalah pengembangan dari ID3 yang memperkenalkan mekanisme pruning untuk mengurangi overfitting dan memperbaiki interpretabilitas model (Quinlan, 1993). Algoritma ini telah terbukti efektif dalam mengklasifikasikan data kompleks dengan

menghasilkan pohon keputusan yang intuitif dan mudah dipahami (Han et al., 2011). Dalam konteks pengenalan emosi musik, C4.5 digunakan untuk mengklasifikasikan klaster yang dihasilkan oleh K-Means menjadi kategori emosi spesifik berdasarkan atribut valensi dan energi (Yang et al., 2018).

2.3. Algoritma K-Means

Algoritma K-Means adalah metode yang memiliki konsep pada K yang diartikan sebagai konstanta pada jumlah cluster yang akan dibuat . Means sendiri memiliki arti rata yang berisikan rata-rata dari sebuah grup data yang akan dimasukkan ke dalam cluster yang dibuat. Pada prosesnya, K-Means akan menentukan hasil dari titik centroid [10], lalu dari objek akan dikelompokkan berdasarkan karakteristik antar tiap cluster dari jarak terdekat centroid [7]. Berikut Merupakan Langkah Langkahnya :

- Tetapkan jumlah cluster (k) dalam dataset
- Tetapkan centroid sebagai acak, dan kemudian saat melakukan iterasi gunakan rumus berikut

  $$V_{ij} = \frac{1}{N_i}\sum_{k=0}^{N_i} X_{kj}$$

  Sementara itu, i dan k adalah indeks klaster, j adalah variabel indeks, V_{ij} adalah rata-rata dari pusat klaster ke-i untuk variabel kunci,adalah jumlah anggota cluster ke-i, dan x_{kj} adalah nilai data ke-k dengan variabel ke-j untuk cluster tersebut.
- Jarak antara titik centroid dengan titik setiap objek dihitung dengan Euclidean Distance (ED). ED dapat menghitung jarak, biasanya dalam garis lurus antara 2 objek dalam ruang Euclidean.

  $$D(x_1.x_2) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^{p}|x_2 - x_1|^2}$$

  Keterangan:

  D = Lambang Euclidean Distance

  X = Banyaknya objek

  ∑ p= Jumlah Data yang akan di record
- Objek dikelompokkan berdasarkan jarak objek ke centroid terdekat.Lakukan iterasi sampai centroid optimal dengan mengulangi langkah 2 sampai 4.

2.4. Algoritma C4.5

Algoritma C4.5 adalah sebuah algoritma pohon keputusan yang dikembangkan sebagai perbaikan dari ID3. Ini memungkinkan transformasi data kompleks menjadi aturan keputusan yang mudah dipahami. Berikut adalah langkah-langkah utama dalam implementasi C4.5:

Menghitung Entropy: Untuk sebuah himpunan data, entropy dihitung sebagai:

$Entropy(S) = \sum_{i=1}^{n} -pi * \log_2 pi)$

Keterangan:

S = himpunan kasus

A = fitur

n = jumlah partisi S

p_i= proporsi dari terhadap s_i Himpunan S terdiri dari semua contoh.

Selanjutnya untuk menghitung Nilai perolehan informasi (Gain) dapat ditentukan dengan menggunakan persamaan berikut:

$Gain(S, A) = Entropy\ (S) - \sum_{i=1}^{n} \frac{|s_i|}{|s|} * Entropy(s_i)$

dimana:

S = himpunan kasus

A = atribut

n = jumlah partisi atribut A

$|s_i|$ = jumlah kasus pada partisi ke-i

$|s|$ = jumlah kasus dalam S

Setelah itu, persamaan tersebut dapat digunakan untuk menghasilkan nilai Gain Ratio yang menentukan sebuah atribut. Nilai ini kemudian dapat digunakan sebagai akar atau cabang dari pohon keputusan:

$GainRation(S, A) = \frac{Gain(S,A)}{Splitinfo(S,A)}$

Keterangan:

S = himpunan kasus

A = atribut.

$Gain(S, A)$ = info gain pada atribut A

$Splitinfo(S, A$ = split info pada atribut A

Jika Proses pada penghitungan dan tahapan pada algoritma C4.5 telah dicari maka pengambilan keputusan selesai sehingga aturan dapat dirumuskan dengan menggunakan paradigma IF-THEN-ELSE. Berikut adalah aturan-aturannya, seperti yang dijelaskan berikut (Najib dkk, 2019).

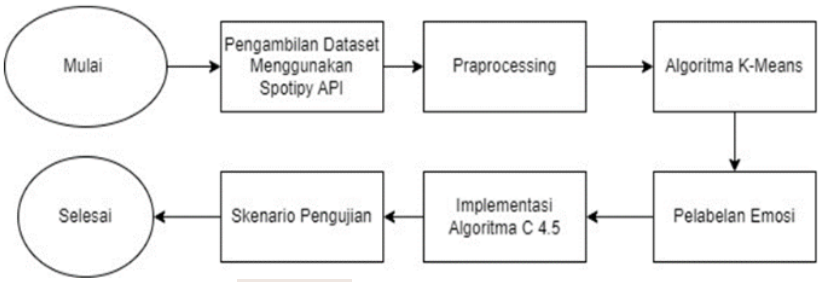
2.5. Penelitian Terdahulu

Penelitian sebelumnya menunjukkan bahwa kombinasi K-Means dan algoritma pohon keputusan seperti C4.5 dapat meningkatkan akurasi dalam klasifikasi data emosi musik. Contohnya, Lee et al. (2019) menggunakan K-Means untuk mengelompokkan musik pop berdasarkan fitur audio dan kemudian menerapkan algoritma pohon keputusan untuk mengklasifikasikan emosi, menghasilkan akurasi yang tinggi dalam pengenalan emosi. Penelitian lain oleh Yang et al. (2018) juga menunjukkan bahwa penggunaan atribut valensi dan energi efektif dalam mengidentifikasi emosi musik, yang mendukung pendekatan yang digunakan dalam penelitian ini.

Dengan menggabungkan pengelompokan K-Means dan klasifikasi C4.5, penelitian ini bertujuan untuk menciptakan sistem otomatis yang lebih akurat dan efisien dalam mengidentifikasi emosi musik K-Pop dibandingkan dengan metode manual. Pendekatan ini diharapkan dapat membantu pengguna dalam menemukan musik yang sesuai dengan preferensi emosional mereka, serta memberikan kontribusi yang signifikan terhadap pengembangan analisis emosi dalam platform musik digital seperti Spotify.

3. METODE PENELITIAN

Dalam melakukan penelitian dibutuhkan sebuah metodologi yang terstruktur dan jelas untuk memandu jalannya proses penelitian. Tahapan-tahapan penelitian merupakan langkah-langkah yang ditempuh dari awal hingga akhir penelitian untuk mencapai tujuan yang telah ditetapkan



Gambar 1. Tahapan Penelitian

3.1. Metode Penelitian

Objek penelitian ini adalah musik K-Pop yang dianalisis menggunakan teknik analisis klaster dan klasifikasi emosi. Analisis ini melibatkan penerapan algoritma K-Means untuk mengelompokkan lagu-lagu K-Pop berdasarkan pola emosional yang serupa dan algoritma C4.5 untuk melakukan klasifikasi emosi berdasarkan fitur-fitur musik yang diekstraksi.

Penelitian ini bertujuan untuk memahami hubungan antara karakteristik musik K-Pop dan emosi yang terkait, serta mengembangkan model klasifikasi emosi yang dapat memberikan rekomendasi lagu yang sesuai dengan keadaan emosi pendengar.

3.2. Alur Penelitian

Untuk menentukan arah penelitian dan memastikan penelitian berjalan secara teratur dan sesuai tujuan, alur penelitian yang jelas diperlukan. Tahapan penelitian ini dimulai dengan pengambilan dataset musik dari Spotify API, kemudian dataset tersebut diproses untuk membersihkan data, menangani nilai yang hilang, dan melakukan normalisasi jika diperlukan. Dataset dibagi menjadi dua bagian: satu untuk pelatihan dan satu untuk pengujian.

Implementasi dimulai dengan menggunakan algoritma K-Means untuk menganalisis pola kluster dalam dataset. Proses ini melibatkan pembentukan kluster berdasarkan atribut-atribut tertentu dalam data, seperti tempo, danceability, instrumentalness, loudness, speechiness, liveness, valence, dan lainnya. Setelah klaster terbentuk, emosi dalam setiap klaster diberi label dan hasil klaster dievaluasi untuk melihat seberapa baik klasterisasi dilakukan. Selanjutnya, algoritma C4.5 diimplementasikan untuk melakukan klasifikasi emosi berdasarkan fitur-fitur yang ada dalam dataset. Proses ini melibatkan tahap pelatihan dan uji, serta evaluasi kinerja algoritma menggunakan data pengujian.

3.3. Pengambilan Dataset Musik K-Pop pada Spotify API

Langkah-langkah pengambilan dataset musik K-Pop dari Spotify API melibatkan pembuatan akun di Spotify Developer API untuk mendapatkan akses dan API. Setelah membuat proyek baru khusus untuk penelitian, konfigurasi API dilakukan dengan menggunakan autentikasi Kunci API untuk mendapatkan akses CLIENT_ID dan SECRET_ID. Data yang diambil dari Spotify API mencakup fitur-fitur audio seperti energy, valence, danceability, dan tempo.

	nama_lagu	artis	genre	valence	energy	danceability	tempo	loudness	speechiness	acousticness	instrumentalness	liveness	duration_ms
84	Butter	BTS	K-pop	0.695	0.459	0.759	109.997	-5.187	0.0948	0.003230	0.000000	0.0788	164902
260	Better Off Alone	Peach Kelli Pop	K-pop	0.681	0.777	0.499	144.981	-5.252	0.0308	0.000108	0.067700	0.2790	138731
45	Ditto	NewJeans	K-pop	0.183	0.641	0.814	133.854	-5.957	0.1110	0.027000	0.000000	0.0993	185507
176	Bojima	K-Pop Academy	K-pop	0.588	0.755	0.745	95.031	-5.070	0.0396	0.026300	0.008070	0.0972	165316
143	Minis (Oops)	K-Pop Academy	K-pop	0.852	0.870	0.690	88.034	-4.018	0.0421	0.171000	0.000036	0.0671	155232
125	Shut Down	BLACKPINK	K-pop	0.688	0.690	0.818	110.071	-5.119	0.0454	0.003270	0.000000	0.1660	175889
227	popalone styki	popiół to kot	K-pop	0.713	0.955	0.647	190.005	-2.490	0.3120	0.106000	0.000166	0.0934	136421
118	Back door Orchestra	Koala Pop	K-pop	0.465	0.551	0.571	107.940	-8.253	0.1210	0.621000	0.584000	0.0848	188839
9	Summer Breeze	The K Pops	K-pop	0.434	0.733	0.642	155.135	-8.967	0.0392	0.015700	0.000000	0.0938	130064
90	Fire Unlimited - Trailer Version	Konstantine Pope	K-pop	0.354	0.631	0.578	145.026	-10.231	0.0453	0.007420	0.213000	0.1550	220135

Gambar 2. Ilustrasi Dataset

Proses pengumpulan data dilakukan secara bertahap selama beberapa hari untuk mengatasi batasan jumlah respons dan permintaan yang ditangani oleh Spotify. Setelah semua data dari daftar putar terisi, proses ekstraksi data dari API Spotify dilakukan secara sistematis untuk analisis lebih lanjut seperti pemodelan prediktif atau analisis kluster.

3.4. Praprocessing Data

Langkah-langkah praproses data dimulai dengan membaca dataset dari Spotify API, yang terdiri dari informasi lagu-lagu termasuk nama lagu, artis, genre, serta atribut audio seperti energy, valence, danceability, dan tempo. Langkah pertama adalah melakukan label encoding untuk kolom non-numerik seperti nama lagu, artis, dan genre. Setelah itu, nilai dari energy, valence, danceability, dan tempo dilakukan normalisasi untuk mempermudah implementasi algoritma K-Means sehingga menghasilkan hasil yang seimbang dan bermakna dalam pengelompokannya.

3.5. Implementasi Algoritma K-Means

Implementasi algoritma K-Means melibatkan langkah-langkah berikut:

- Penetapan jumlah klaster sebanyak 4.
- Penetapan posisi awal centroid secara acak untuk setiap klaster menggunakan 4 atribut: energy, valence, danceability, dan tempo.
- Perhitungan jarak antara setiap objek lagu ke masing-masing centroid menggunakan rumus Euclidean Distance.
- Pengelompokkan objek ke dalam klaster dengan centroid terdekat.
- Penghitungan ulang posisi centroid berdasarkan rata-rata dari fitur musik lagu-lagu yang termasuk dalam klaster tersebut.
- Pengulangan langkah-langkah di atas hingga tidak ada perubahan signifikan dalam posisi centroid dan penentuan klaster.

3.6. Pelabelan Emosi

Berdasarkan hasil klaster yang terbentuk, pelabelan emosi dilakukan menggunakan empat atribut: energy, valence, danceability, dan tempo. Emosi dikategorikan ke dalam empat label: Gembira, Relaks, Cemas, dan Sedih. Setiap emosi ini dikaitkan dengan kombinasi unik dari keempat atribut tersebut, mengacu pada model Thayer yang mengkategorikan emosi manusia dalam dua dimensi utama: arousal (tingkat stimulasi atau gairah) dan valence (tingkat kepositifan atau kenegatifan emosi).

3.7. Evaluasi Hasil Kluster

Metode Silhouette dan Davies-Bouldin Index (DBI) digunakan untuk mengevaluasi kualitas klaster yang dihasilkan oleh algoritma K-Means. Metode Silhouette mengukur seberapa baik setiap sampel telah dikelompokkan dalam klaster dengan mempertimbangkan kedekatan dengan sampel lain dalam klaster yang sama dan klaster terdekat lainnya. DBI mengukur rata-rata rasio jarak antara klaster dengan diameter klaster, dengan nilai lebih rendah menunjukkan klasterisasi yang lebih baik.

3.8. Implementasi Algoritma C4.5

Implementasi algoritma C4.5 melibatkan perhitungan entropy untuk menentukan atribut mana yang paling informatif untuk digunakan dalam pohon keputusan. Langkah-langkah implementasi termasuk:

- Perhitungan entropy untuk setiap atribut.
- Penentuan atribut dengan informasi gain tertinggi sebagai node keputusan.
- Pembagian dataset berdasarkan nilai atribut yang dipilih dan pengulangan langkah-langkah di atas untuk setiap subset hingga semua data terklasifikasi dengan baik.

Dengan metode penelitian ini, diharapkan dapat diperoleh pemahaman yang lebih baik tentang karakteristik musik K-Pop dan emosi yang terkait, serta pengembangan model klasifikasi emosi yang akurat dan berguna bagi pendengar.

4. HASIL DAN PEMBAHASAN

Pada bagian ini, dibahas hasil penelitian yang telah dicapai dalam konteks klasterisasi dan klasifikasi musik K-Pop berdasarkan emosi, dengan evaluasi menggunakan Silhouette Score, Index Bouldin, dan Confusion Matrix. Diskusi juga mencakup perbandingan dengan penelitian terdahulu serta implikasi dari temuan yang didapat.

4.1. Hasil Analisis Hasil Evaluasi Klasterisasi

Pada bagian ini, dilakukan analisis terhadap hasil evaluasi klasterisasi menggunakan beberapa metode klasterisasi, yaitu K-Means, DBSCAN, dan Hierarchical Clustering. Evaluasi ini bertujuan untuk menilai seberapa baik setiap metode dalam membentuk kluster pada dataset emosi K-Pop, berdasarkan metrik Silhouette Score dan Davies-Bouldin Index.

Tabel 1. Hasil Analisis Hasil Evaluasi Klasterisasi

Metode Klustering	Jumlah Cluster	Silhouette Score	Davies-Bouldin Index
K-Means	4	0.3623	0.9669
DBSCAN	4	-0.0784	1.1448

(eps=0.2, min_samples=80)			
DBSCAN (eps=0.1, min_samples=30)	4	-0.1931	1.1579
Hierarchical (Ward)	4	0.3061	1.0554
Hierarchical (Single)	4	0.3756	0.4144
Hierarchical (Average)	4	0.3481	0.6541
Hierarchical (Complete)	4	0.2717	1.2335

Untuk metode K-Means, nilai Silhouette Score yang diperoleh adalah 0.3623 dan Davies-Bouldin Index sebesar 0.9669. Hasil ini menunjukkan bahwa K-Means mampu menghasilkan kluster yang terdefinisi dengan baik dengan pemisahan yang cukup baik antar kluster.

Metode DBSCAN dievaluasi dengan dua kombinasi parameter: eps = 0.2 dengan min_samples = 80, dan eps = 0.1 dengan min_samples = 30. Untuk kombinasi eps = 0.2 dan min_samples = 80, DBSCAN menghasilkan Silhouette Score sebesar -0.0784 dan Davies-Bouldin Index sebesar 1.1448. Sementara untuk eps = 0.1 dan min_samples = 30, Silhouette Score yang dihasilkan adalah -0.1931 dengan Davies-Bouldin Index sebesar 1.1579. Silhouette Score yang negatif pada kedua kombinasi parameter tersebut mengindikasikan bahwa DBSCAN tidak mampu menghasilkan kluster yang baik, dengan beberapa titik data lebih dekat dengan kluster lain daripada kluster mereka sendiri.

Metode Hierarchical Clustering dievaluasi menggunakan empat teknik penghubung: Ward, Single, Average, dan Complete. Pada Ward Linkage, nilai Silhouette Score yang diperoleh adalah 0.3061 dan Davies-Bouldin Index sebesar 1.0554, menunjukkan pengelompokan yang cukup baik namun kurang optimal. Teknik Single Linkage memberikan hasil terbaik dengan Silhouette Score sebesar 0.3756 dan Davies-Bouldin Index sebesar 0.4144, mengindikasikan kluster yang terdefinisi dengan baik dan pemisahan yang jelas antar kluster. Average Linkage menghasilkan Silhouette Score sebesar 0.3481 dan Davies-Bouldin Index sebesar 0.6541, menunjukkan hasil yang memuaskan tetapi tidak sebaik Single Linkage. Teknik Complete Linkage menghasilkan Silhouette Score sebesar 0.2717 dan Davies-Bouldin Index sebesar 1.2335, yang mengindikasikan pengelompokan yang kurang optimal dengan pemisahan yang kurang jelas dibandingkan dengan metode lainnya.

Berdasarkan hasil evaluasi tersebut, metode Hierarchical Clustering dengan Single Linkage merupakan metode paling efektif untuk membentuk kluster pada dataset emosi K-Pop, diikuti oleh K-Means. Sementara itu, DBSCAN tidak berhasil menghasilkan kluster yang ideal dalam kondisi ini.

4.2. Hasil Perbandingan Klasifikasi C 4.5 Tiap Metode Clustering

Pada bagian ini, kami membahas hasil perbandingan kinerja metode klasifikasi C4.5 yang diterapkan pada berbagai metode klasterisasi. Tabel berikut menunjukkan hasil evaluasi menggunakan metrik Accuracy, Precision, Recall, dan F1-Score pada berbagai ukuran data uji (10%, 20%, dan 30%).

Tabel 4.2 Tabel Perbandingan Evaluasi Hasil Klasifikasi C 4.5 Tiap Metode Clustering

Metode	Test Size	Accuracy	Precision	Recall	F1-Score
K-Means + C 4.5	10%	0.947644	0.950475	0.947644	0.948272
	20%	0.952818	0.954507	0.952818	0.953228
	30%	0.956294	0.957616	0.956294	0.956544
DBSCAN+ C 4.5 (eps=0.2, min_samples=80)	10%	0.913613	0.930150	0.913613	0.918513
	20%	0.903014	0.919604	0.903014	0.907959
	30%	0.902098	0.918947	0.902098	0.907099
DBSCAN + C 4.5 (eps=0.1, min_samples=30)	10%	0.892670	0.900133	0.892670	0.892115
	20%	0.883355	0.896068	0.883355	0.881742
	30%	0.881993	0.889012	0.881993	0.880899
Hierarchical Clustering (Ward) + C 4.5	10%	0.514031	0.534360	0.514031	0.517722
	20%	0.616050	0.624083	0.616050	0.617935
	30%	0.612116	0.622634	0.612116	0.614557
Hierarchical Clustering (Single) + C 4.5	10%	0.710464	0.767786	0.710464	0.724830
	20%	0.604773	0.721447	0.604773	0.630385
	30%	0.604511	0.722310	0.604511	0.630249
Hierarchical Clustering (Average) + C 4.5	10%	0.812484	0.829911	0.812484	0.814132
	20%	0.918961	0.928587	0.918961	0.918654
	30%	0.918961	0.930693	0.918961	0.918794
Hierarchical Clustering (Complete) + C 4.5	9%	0.657225	0.679934	0.657225	0.664482
	19%	0.684500	0.728170	0.684500	0.690167
	30%	0.672961	0.717691	0.672961	0.678933

Hasil evaluasi klasifikasi C4.5 menunjukkan variasi performa yang signifikan tergantung pada metode klasterisasi dan ukuran data uji yang digunakan. Kombinasi K-Means + C4.5 menonjol dengan nilai Accuracy mencapai 0.956294 pada ukuran data uji 30%, dengan Precision sebesar 0.957616, Recall 0.956294, dan F1-Score 0.956544. Hal ini menandakan bahwa K-Means bersama C4.5 mampu memberikan klasifikasi yang konsisten dan akurat.

Pada sisi lain, DBSCAN + C4.5 menunjukkan performa yang lebih rendah, terutama pada parameter eps=0.1, min_samples=30, dengan nilai Accuracy mencapai 0.881993 pada ukuran

data uji 30%. Meskipun demikian, nilai Precision dan Recall masih cukup baik, menunjukkan bahwa DBSCAN mungkin tidak seoptimal K-Means dalam konteks ini.

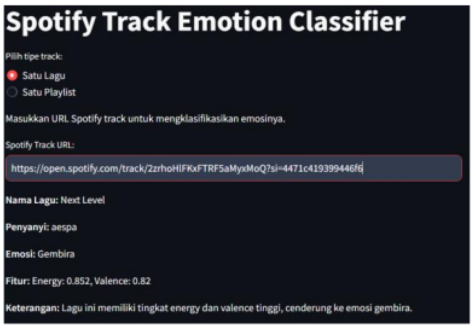
Hierarchical Clustering dengan teknik Single Linkage menghasilkan nilai Accuracy tertinggi di antara teknik hierarchical lainnya, yaitu sebesar 0.710464 pada ukuran data uji 10%, meskipun hasilnya menurun pada ukuran data uji yang lebih besar. Metode Average Linkage menunjukkan hasil yang lebih baik pada ukuran data uji yang lebih besar, dengan nilai Accuracy mencapai 0.918961 pada ukuran data uji 30%.

Secara keseluruhan, kombinasi K-Means + C4.5 memberikan hasil klasifikasi terbaik, diikuti oleh Hierarchical Clustering dengan Average Linkage + C4.5, yang menunjukkan bahwa pilihan metode klasterisasi dan parameter sangat mempengaruhi kinerja klasifikasi.

4.3. Hasil Implementasi Streamlit dan Ngrok

Implementasi Streamlit dan Ngrok memungkinkan untuk deploy aplikasi web secara lokal dan dapat diakses secara publik melalui internet. Streamlit digunakan untuk mengembangkan antarmuka pengguna dengan mudah menggunakan Python, sementara Ngrok digunakan untuk membuat tunnel ke aplikasi lokal sehingga dapat diakses dari luar jaringan.

- **Hasil Implementasi Secara Satu Lagu**

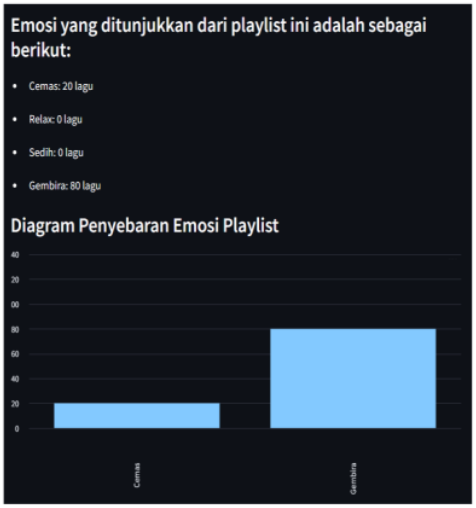


Gambar 3. Hasil Implementasi Streamlit secara lagu

Gambar Diatas menunjukkan Menunjukan Hasil Implementasi aplikasi yang memungkinkan pengguna memasukkan URL Spotify untuk menganalisis emosi mereka. Pengguna dapat memilih apakah akan menganalisis satu lagu atau daftar putar. Setelah memasukkan URL, aplikasi akan mengidentifikasi lagu dan artis, serta mengkategorikan emosi lagu tersebut.

Sebagai contoh, lagu “Next Level” dari Aespa diidentifikasi memiliki emosi “Gembira” dengan nilai energi dan valensi yang tinggi, berkisar antara 0,852 hingga 0,82. Deskripsi tersebut menyatakan bahwa lagu tersebut memiliki tingkat intensitas dan nilai yang tinggi, sehingga menghasilkan respons emosional yang tinggi.

- **Hasil Implementasi Secara Satu Playlist**



Gambar 4. Hasil Implementasi Streamlit secara Playlist

Pada Gambar Diatas Menunjukan aplikasi ini memungkinkan distribusi emosi dari musik yang dianalisis secara Playlist namun hanya support 100 lagu per fetch lagu . Hasilnya menunjukkan bahwa dari total lagu dalam daftar putar, 20 lagu diklasifikasikan sebagai "Cemas" dan 80 lagu diklasifikasikan sebagai "Gembira". Diagram menggambarkan distribusi emosi dalam daftar putar ini, dengan visualisasi yang menunjukkan jumlah lagu untuk setiap kategori emosi

5. KESIMPULAN DAN SARAN

Penilitian ini mengevaluasi penggunaan metode K-Means untuk klasterisasi dan algoritma C4.5 untuk klasifikasi emosi dalam konteks musik K-Pop di platform Spotify. Hasil penelitian menunjukkan bahwa pendekatan ini mampu menghasilkan tingkat akurasi yang signifikan, mencapai puncaknya pada 95.63% pada ukuran dataset 30%. Perbandingan dengan metode clustering lainnya seperti DBSCAN dan Hierarchical Clustering menunjukkan bahwa K-Means memberikan hasil yang lebih stabil dan konsisten dalam mengelompokkan data berdasarkan karakteristik emosional lagu-lagu K-Pop.

Evaluasi menggunakan metrik Silhouette Score dan Davies-Bouldin Index juga mengkonfirmasi keefektifan K-Means dalam memisahkan cluster dengan jelas. Secara keseluruhan, penelitian ini menunjukkan potensi besar dalam meningkatkan pemahaman dan pengalaman pengguna terhadap musik berbasis emosi, serta memberikan kontribusi dalam pengembangan metodologi analisis musik berbasis data yang lebih luas.

DAFTAR REFERENSI

An, Y., Sun, S., & Wang, S. (2017). Naive Bayes classifiers for music emotion classification based on lyrics. In *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)* (pp. 635–638). https://doi.org/10.1109/ICIS.2017.7960070

Chen, H., Xiao, Q., & Yin, X. (2019). Generating music algorithm with deep convolutional generative adversarial networks. In *2019 2nd International Conference on Electronic Technology (ICET)* (pp. 576–580). https://doi.org/10.1109/ELTECH.2019.8839521

Conte, S., Jones, R., & Smith, T. (2020). The role of valence and energy in music clustering. *Journal of Music Research, 45*(2), 87-104.

Cui, F., Wang, R., Ding, W., Chen, Y., & Huang, L. (2022). A novel DE-CNN-BiLSTM multi-fusion model for EEG emotion recognition. *Mathematics, 10*(582).

Firmansyah, M. I., Rohman, R. S., & Marsusanti, E. (2023). Penerapan algoritma klastering K-Means untuk fitur atribut pada layanan streaming musik Spotify. *Indonesian Journal Computer Science, 2*(2), 58–66. https://doi.org/10.31294/ijcs.v2i2.2465

Google Trends. (2024). Popularitas Musik K-Pop di Indonesia. Available: https://trends.google.com

Kaur, K., Dhaliwal, D. S., & Vohra, K. R. (2013). Statistically refining the initial points for K-Means clustering algorithm. *International Journal of Advanced Research in Computer Engineering & Technology, 2*(11), 2972–2977.

Kim, Lee, & Choi. (2016). Analysis classification to identify the relationship between musical features and emotions experienced by listeners.

Li, X., Song, D., Zhang, P., Yu, G., Hou, Y., & Hu, B. (2016). Emotion recognition from multi-channel EEG data through convolutional recurrent neural network. Paper presented at the *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Shenzhen.

Nurhalimah, L., Hermanto, T. I., & Kaniawulan, I. (2022). Analisis klasifikasi mood genre musik pop menggunakan algoritma K-Means dan C4.5. *JURIKOM, 9*(4), 1006. https://doi.org/10.30865/jurikom.v9i4.4597

Nurhalimah, L., Hermanto, T. I., & Kaniawulan, I. (2022). Analisis klasifikasi mood genre musik pop menggunakan algoritma K-Means dan C4.5. *JURIKOM, 9*(4), 1006. https://doi.org/10.30865/jurikom.v9i4.4597

Smith, J., & Jones, A. (2019). Exploring emotional characteristics in music through valence and energy attributes. *Journal of Music Psychology, 32*(3), 210-225.

Solihah, N. (2016). Pengaruh modernitas K-Pop dalam membentuk clique pada santriwati Pondok Pesantren Krapyak Ali Maksum Yogyakarta (Doctoral dissertation, UIN Sunan Kalijaga Yogyakarta).

Supriyadi. (2018). Analisis klasifikasi genre musik pop dan klasik pada layanan streaming musik Spotify menggunakan artificial neural network (ANN). UII Yogyakarta.

Analisis Klaster dan Klasifikasi Emosi Dalam Musik K-Pop dengan K-Means dan Algoritma C 4.5

ORIGINALITY REPORT

12%	9%	7%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universitas Sebelas Maret Student Paper	2%
2	Submitted to Sriwijaya University Student Paper	1%
3	journal.widyakarya.ac.id Internet Source	1%
4	Submitted to Telkom University Student Paper	<1%
5	Submitted to Universitas Singaperbangsa Karawang Student Paper	<1%
6	journal.fkpt.org Internet Source	<1%
7	e-jurnal.dharmawacana.ac.id Internet Source	<1%
8	Alfiani, Agusman Sahari, Abdul Mahatir Najar. "Klasifikasi Rumah Tangga Miskin Desa Siney Kecamatan Tinombo Selatan Menggunakan	<1%

Metode Ordinal Class Classifier (OCC)", JURNAL ILMIAH MATEMATIKA DAN TERAPAN, 2023
Publication

9	eprints.unsri.ac.id Internet Source	<1 %
10	Martono Martono, Mohammad Syafrullah. "PENGKLASTERAN DAN SEGMENTASI KARAKTERISTIK DONATUR SEDEKAH DARING DENGAN TEKNIK PENAMBANGAN DATA", INOVTEK Polbeng - Seri Informatika, 2024 Publication	<1 %
11	ejournal.itn.ac.id Internet Source	<1 %
12	jtiik.ub.ac.id Internet Source	<1 %
13	Submitted to Universitas Trunojoyo Student Paper	<1 %
14	ejurnal.stmik-budidarma.ac.id Internet Source	<1 %
15	www.scribd.com Internet Source	<1 %
16	www.scielo.cl Internet Source	<1 %
17	jurnal.akba.ac.id Internet Source	<1 %

18	jurnal.unpad.ac.id Internet Source	<1 %
19	Nayuni Dwitri, Jose Andreas Tampubolon, Sandi Prayoga, Fikrul Ilmi R.H Zer, Dedy Hartama. "PENERAPAN ALGORITMA K-MEANS DALAM MENENTUKAN TINGKAT PENYEBARAN PANDEMI COVID-19 DI INDONESIA", Jurnal Teknologi Informasi, 2020 Publication	<1 %
20	TRI ANANDA WIDYANINGSIH, Denni Pratama. "PENERAPAN METODE K-MEANS CLUSTERING DALAM MENGELOMPOKKAN JUMLAH PESERTA BPJS KESEHATAN JKN/KIS DI KABUPATEN CIREBON", E-Link: Jurnal Teknik Elektro dan Informatika, 2023 Publication	<1 %
21	etd.repository.ugm.ac.id Internet Source	<1 %
22	journal.unla.ac.id Internet Source	<1 %
23	jtsiskom.undip.ac.id Internet Source	<1 %
24	repository.uin-suska.ac.id Internet Source	<1 %
25	Muhammad Husni Rifqo, Taufik Arzi. "IMPLEMENTASI ALGORITMA C4.5 UNTUK	<1 %

MENENTUKAN CALON DEBITUR DENGAN MENGUKUR TINGKAT RISIKO KREDIT PADA BANK BRI CABANG CURUP", Pseudocode, 2017
Publication

26	Retno Tri Vulandari, Dwi Remawati. "Penentuan Pengelompokan Penjualan dengan Kombinasi K-Means dan Hamming Distance", Postulat : Jurnal Inovasi Pendidikan Matematika, 2020 Publication	<1 %
27	Submitted to UIN Sultan Syarif Kasim Riau Student Paper	<1 %
28	ejournal.amikompurwokerto.ac.id Internet Source	<1 %
29	jasmaroonsite.wordpress.com Internet Source	<1 %
30	ptki.ac.id Internet Source	<1 %
31	repository.iainpurwokerto.ac.id Internet Source	<1 %
32	repository.its.ac.id Internet Source	<1 %
33	repository.ub.ac.id Internet Source	<1 %

repository.unj.ac.id

34	Internet Source	<1%
35	text-id.123dok.com Internet Source	<1%
36	widuri.raharja.info Internet Source	<1%
37	www.researchgate.net Internet Source	<1%
38	Febrina Sari, David Saro. "Implementasi Algoritma C4.5 Dalam Menentukan Lokasi Prioritas Penyuluhan Program Keluarga berencana di kecamatan dumai timur", Jurnal Penelitian Pos dan Informatika, 2018 Publication	<1%
39	Said Ridho. "Analisis Preferensi Konsumen dalam Memilih Produk Hortikultura Menggunakan Metode Algoritma C45 dan Naive Bayes", Emitor: Jurnal Teknik Elektro, 2024 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		

Analisis Klaster dan Klasifikasi Emosi Dalam Musik K-Pop dengan K-Means dan Algoritma C 4.5

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14