KLASTERISASI DATA LAGU TERPOPULER *SPOTIFY* 2023 BERDASARKAN SUASANA HATI MENGGUNAKAN ALGORITMA *K-MEANS*

Dwi Nuriska, Bambang Irawan, Agus Bahtiar, Arif Rinaldi Dikananda
Teknik Informatika, STMIK IKMI Cirebon
Jalan Perjuangan 10B Majasem, Karyamulya, Kesambi, Kota Cirebon, Indonesia
dwinuriska4@gmail.com

ABSTRAK

Spotify sebagai platfrom *streaming* musik, menampilkan berbagai fitur yang beragam dan secara terus-menerus diperbarui untuk mencerminkan perkembangan dalam dunia musik. Permasalahan penelitian muncul dari beragamnya preferensi pengguna dan tren mendengarkan sebuah lagu atau musik mengakibatkan kompleksitas dalam pemahaman dan pengelompokan musik. Tujuan penelitian ini adalah menghasilkan kelompok lagu yang lebih terfokus berdasarkan suasana hati, memungkinkan pengguna untuk lebih mudah menemukan lagu yang sesuai dengan *mood* atau suasana hati mereka. Dalam penelitian ini akan mengelompokkan data lagu-lagu terpopuler Spotify 2023 berdasarkan empat kategori suasana hati model *Thayer's (Angry, Sad, Relax, dan Happy)* dengan menggunakan Algoritma *K-Means Clustering*. Dengan pendekatan *Knowledge Discovery in Database* (KDD). Atribut-atribut audio seperti tempo *(bpm), danceability*, *valence*, dan *energy* digunakan dalam analisis ini. Hasil penelitian diperoleh nilai *Davies-Bouldin Index* (DBI) terkecil adalah 0,299 dengan jumlah K = 3, di mana *cluster* 0 merupakan suasana hati *happy* dengan 361 anggota lagu, *cluster* 1 merupakan suasana hati *relax* dengan 329 anggota lagu, dan *cluster* 2 merupakan suasana hati *sad* dengan 252 anggota lagu. Tidak terdapat *cluster* yang secara jelas menggambarkan suasana hati marah. Dengan distribusi anggota yang merata di setiap *cluster*, maka tidak ada suasana hati yang mendominasi dari hasil pengelompokan lagu.

Kata kunci : Klasterisasi, *K-Means*, Lagu terpopuler *Spotify* 2023, Suasana Hati

1. PENDAHULUAN

Penyedia layanan untuk mengakses dan mendengarkan musik digital melalui jaringan internet saat ini terdiri dari banyak pilihan. Beragamnya opsi yang tersedia ini memberikan kemudahan serta kebebasan kepada masyarakat dalam memilih layanan yang sesuai dengan preferensi dan kebutuhan mereka. Kegiatan mendengarkan musik telah menjadi sebuah kegiatan rutin yang penting dan tidak boleh diabaikan [1].

Musik adalah serangkaian nada yang diorganisasikan secara sistematis sehingga menghasilkan suatu kesatuan komposisi suara yang memiliki irama dan harmoni. Ketika musik dimainkan dengan komposisi terpadu, dampaknya dapat dirasakan pada aspek emosional dan suasana hati [2]. Musik memiliki dampak yang penting pada aspek emosional, di mana musik menghasilkan kesan yang menyenangkan dan menggetarkan dapat meningkatkan semangat. Selain itu, tempo dalam musik juga memiliki pengaruh yang signifikan terhadap emosi dalam berkomunikasi. Tempo musik yang lebih cepat dapat meningkatkan semangat dan rasa senang [3]. Musik memiliki potensi untuk menyediakan sumber energi positif ketika kita menghadapi kelelahan atau kesulitan dalam menanggapi suatu masalah. Musik berfungsi sebagai pembangkit suasana hati yang efektif ketika kita merasa jenuh, baik melalui mendengarkan dan menikmati musik, maupun melalui aktivitas seperti bermain alat musik dan menyanyi secara langsung keduanya mampu mengubah kondisi mental dan perasaan yang sedang kita alami [4]. Oleh karena itu, musik atau lagu dapat dianggap sebagai alat untuk mendukung suasana hati pengguna.

Salah satu *platform streaming* musik yang populer adalah *Spotify*. *Spotify* merupakan *platform* layanan *streaming* musik yang menyediakan fasilitas mendengarkan musik secara *online* bagi pengguna. *Spotify* diperkenalkan pertama kali pada tanggal 1 April 2006 di kota Stockholm, Swedia, sebagai langkah awal dalam memberikan *platform* layanan musik digital. Sebagai salah satu *platform streaming* musik terkemuka di dunia, *Spotify* memberikan sarana untuk berkomunikasi secara konsisten dengan cara individu menerima informasi [5]. Permasalahan penelitian muncul dari beragamnya preferensi pengguna dan tren mendengarkan sebuah lagu atau musik mengakibatkan kompleksitas dalam pemahaman dan pengelompokan musik. Oleh karena itu penelitian ini penting untuk dilakukan agar permasalahan tersebut bisa terselesaikan.

Sebelumnya, telah banyak penelitian yang berkaitan dengan penerapan Algoritma *K-Means Clustering* dan pengelompokan lagu berdasarkan suasana hati atau *mood* dalam berbagai konteks, seperti penelitian yang dilakukan oleh Rubangiya, Tuti Hartati & Yudhistira Arie Wijaya, yang berjudul "Analisis Data Lalu Lintas Jaringan Di Kantor *Cangehgar Cyber Operation* Center Menggunakan Algoritma *K-Means* " membahas mengenai analisis data lalu lintas jaringan menggunakan Algoritma *K-Means* dengan hasil menemukan protokol seperti TCP, MDNS, dan HTTP sebagai protokol yang paling

umum digunakan [6]. Sebaliknya, pada penelitian kedua yang dilakukan oleh Lia Nurhalimah, Teguh Iman Hermanto, dan Ismi Kaniawulan, yang berjudul "Analisis Prediksi *Mood* Genre Musik Pop Menggunakan Algoritma *K-Means* dan C4.5" membahas mengenai pengelompokan genre musik pop berdasarkan *mood* dengan hasil penelitian menunjukkan bahwa *mood* yang paling dominan pada genre musik pop adalah *cheerful*. Evaluasi dilakukan menggunakan *confusion matrix* dan tingkat akurasi yang diperoleh sebesar 91,9% [7]. Sementara itu, penelitian ketiga dilakukan oleh Fuad Fadlila Surenggana, Arik Aranta dan Fitri Bimantoro, yang berjudul "Klasifikasi *Mood* Musik Menggunakan *K-Nearest Neighbor* Dengan *Mel Frequency Cepstral Coefficients*" yang membahas mengenai klasifikasi *mood* musik menggunakan *K-Nearest Neighbor* dengan *Mel Frequency Cepstral Coefficients* (MFCC). *Dataset* yang digunakan 200 file musik yang dibagi menjadi 4 kelas *mood*. Penelitian mencapai akurasi 85,5% menggunakan KNN dengan nilai k=5 dan metode jarak Manhattan [3].

Penelitian ini bertujuan untuk menentukan jumlah *cluster* optimal dalam pengelompokan lagu-lagu dengan menggunakan *Davies-Bouldin Index* sebagai metode evaluasi. Selain itu, penelitian ini mengarah pada tujuan mengelompokkan lagu-lagu terpopuler *Spotify* tahun 2023 berdasarkan suasana hati model Thayer's (*angry, sad, Relax, happy*) dengan menggunakan algoritma *K-Means*, sekaligus mendalami pada korelasi dominasi suasana hati dalam kelompok lagu dengan pengalaman mendengarkan musik. Manfaat penelitian ini terletak pada kontribusinya terhadap pengisian kesenjangan pengetahuan di domain Informatika, menyajikan wawasan praktis yang dapat memperkaya pemahaman industri musik digital. Fokus pada identifikasi pola dan atribut data dalam lagu-lagu terpopuler di *Spotify* diharapkan memberikan pedoman berharga bagi praktisi industri musik untuk mengoptimalkan manajemen dan distribusi konten musik di *platform* digital. Sehingga, penelitian ini tidak hanya berpotensi meningkatkan pengetahuan akademis, melainkan juga memberikan dampak positif secara praktis dalam konteks industri musik modern.

Metode analisis data yang diterapkan dalam penelitian ini adalah proses *Knowledge Discovery in Database* (KDD) dengan penerapan Algoritma *K-Means Clustering* pada proses data *mining* [6]. KDD merupakan pendekatan yang diterapkan untuk mendapatkan pemahaman dari basis data yang tersedia. Hasil pemahaman ini dapat dimanfaatkan sebagai dasar pengetahuan (*knowledge base*) yang mendukung proses pengambilan keputusan [8]. Analisis hasilnya akan dilakukan secara teliti untuk mengidentifikasi pola pendengaran yang signifikan dan atribut data yang memengaruhi suasana hati lagu. Melalui penerapan KDD dan analisis data yang cermat, diharapkan penelitian ini dapat memberikan wawasan mendalam dan dapat diandalkan terkait preferensi musik pendengar serta dinamika industri musik dalam konteks era digital.

2. TINJAUAN PUSTAKA

2.1. Klasterisasi/*Clustering*

Salah satu teknik dalam *data mining* adalah *Clustering*. *Clustering* merupakan proses pengelompokan sejumlah data atau objek ke dalam kelompok (*Cluster*) sedemikian rupa sehingga setiap kelompok akan berisi data yang memiliki kesamaan yang maksimal dan berbeda dengan objek dalam kelompok lainnya [9].

2.2. *K-Means*

K-Means Clustering adalah suatu metode klasterisasi di mana suatu himpunan data dibagi ke dalam beberapa kelompok. Algoritma *K-Means* pertama kali diajukan oleh *Mac* Queen pada tahun 1967. Metode ini bertujuan untuk mengelompokkan data yang memperlihatkan karakteristik atau pola serupa ke dalam suatu kelompok yang sama, sementara data yang menunjukkan karakteristik atau pola yang berbeda akan dikelompokkan secara terpisah. Proses pengelompokan ini memberikan pemahaman yang lebih dalam terkait informasi atau pengetahuan yang terkandung dalam suatu himpunan data yang luas [10]. Metode ini terutama diterapkan pada atribut *numeric* karena prinsip kerjanya melibatkan pembagian data ke dalam beberapa *cluster* berdasarkan jarak terdekat. Dalam algoritma *K-Means Clustering*, data dikelompokkan berdasarkan karakteristik numeriknya, dengan penentuan *cluster* dilakukan melalui perhitungan jarak terpendek. Metode ini menjadi salah satu algoritma yang sangat umum digunakan dalam konteks pengelompokan data, terutama karena memiliki tahapan yang relatif sederhana [6].

2.3. Lagu Terpopuler *Spotify* 2023

Data lagu terpopuler *Spotify* 2023 merujuk pada informasi mengenai lagu-lagu yang saat ini paling banyak didengarkan oleh pengguna *platform* musik *streaming Spotify*. *Spotify* merupakan *platform* layanan *streaming* musik yang menyediakan fasilitas mendengarkan musik secara *online* bagi pengguna. *Spotify* diperkenalkan pertama kali pada tanggal 1 April 2006 di kota Stockholm, Swedia, sebagai langkah awal dalam memberikan *platform* layanan musik digital. Sebagai salah satu *platform streaming* musik terkemuka di dunia, *Spotify* memberikan sarana untuk berkomunikasi secara konsisten dengan cara individu menerima informasi [5].

2.4. Suasana Hati

Thayer memaparkan bahwa *mood* atau suasana hati adalah rangkaian perasaan yang umumnya bersifat kurang intens dan muncul sebagai respons terhadap situasi dan kondisi yang sedang dialami oleh seseorang [11]. *Thayer* mengusulkan empat kategori *mood* atau suasana hati, yakni *contentment* untuk

menggambarkan suasana hati yang *relax* (santai), *depression* untuk merujuk pada suasana hati sad (sedih), *exuberance* untuk menyatakan suasana hati bahagia (*happy*), dan *anxiety* untuk menggambarkan suasana hati marah (*angry*). Model *Thayer* dipilih karena kesederhanaannya dalam mengelompokkan emosi ke dalam dua dimensi, yaitu *arousal* dan *valence*. Model ini juga diapresiasi karena kemudahan relatifnya dalam diintegrasikan dengan variabel-variabel dari penelitian penggunaan dan fitur audio [3].

2.5. *Davies Bouldin Index* (DBI)

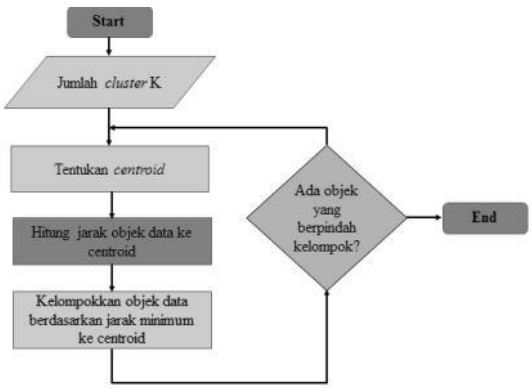
Davies Bouldin Index (DBI) digunakan untuk mengevaluasi dan membandingkan *output* dari algoritma pengelompokan. Pembentukan DBI pertama kali dilakukan pada tahun 1979 oleh *David L. Davies* dan *Donald W. Bouldin. Cluster* yang menggunakan DBI dianggap memiliki skema pengelompokan terbaik ketika nilai DBI-nya rendah. DBI berfungsi sebagai metode untuk mengukur efektivitas atau jumlah *cluster* optimal dalam suatu prosedur pengelompokan, di mana kohesi didefinisikan sebagai penjumlahan kedekatan data dengan titik pusat *cluster* yang mengikutinya. Evaluasi menggunakan DBI terfokus pada skema evaluasi internal *cluster*, di mana jumlah dan kedekatan data dalam hasil *cluster* memberikan informasi tentang seberapa baik atau buruk *cluster* tersebut. DBI digunakan sebagai alat untuk menilai validitas suatu *cluster* dalam metode pengelompokan; semakin rendah atau mendekati 0 nilai DBI, semakin optimal hasil *cluster* tersebut [6].

3. METODE PENELITIAN

3.1. Metode *K-Means Clustering*

Metode *K-Means Clustering* termasuk dalam kategori teknik partisi, di mana objek-objek dibagi atau dipisahkan ke dalam k wilayah yang terpisah. Dalam konteks *K-Means*, setiap objek harus ditempatkan dalam satu kelompok tertentu, namun, dalam suatu tahap proses tertentu, terjadi perpindahan objek dari satu kelompok ke kelompok lainnya.

Proses pengelompokan data menggunakan algoritma *K-Means Clustering* ini diimplementasikan melalui serangkaian langkah-langkah, sebagaimana diperlihatkan dalam Gambar 1.

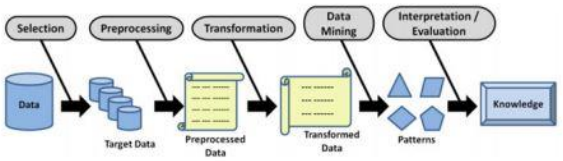


Gambar 1. Alur metode *K-Means* [10]

Proses awal dalam algoritma K-Means Clustering dimulai dengan menentukan jumlah klaster yang diinginkan, kemudian menetapkan pusat klaster (centroid) sebagai representasi titik pusat klaster masing-masing. Selanjutnya, dilakukan perhitungan jarak antara setiap objek data dengan centroid, yang diikuti oleh pengelompokan objek data ke klaster berdasarkan jarak minimum terhadap centroid. Satu iterasi dari proses K-Means Clustering mengacu pada satu siklus lengkap dari langkah-langkah ini. Setelah itu, pusat klaster yang baru dihitung, dan proses perhitungan jarak objek data ke pusat klaster baru dilanjutkan, mirip dengan iterasi sebelumnya. Proses iteratif ini berlanjut hingga hasil pengelompokan pada iterasi tertentu sama dengan iterasi sebelumnya, menandakan konvergensi dan menghasilkan pengelompokan akhir.

3.2. *Knowledge Discovery in Database*

Data yang sudah ada akan dianalisis menggunakan *Knowledge Discovery in Database* (KDD) dengan penerapan Algoritma *K-Means Clustering*, yang terdiri dari enam tahapan pengolahan data terdapat pada Gambar 2.



Gambar 2. Proses KDD [12]

Pada Gambar 2 dapat dijelaskan bahwa masing-masing tahapan *Knowledge Discovery in Database* adalah sebagai berikut :

1. Data
   Dataset awal dibuat dari informasi yang dikumpulkan dari berbagai sumber. Pada tahap ini, data awal yang diperoleh dari berbagai sumber diidentifikasi dan dianggap sebagai potensi sumber pengetahuan yang dapat dianalisis lebih lanjut.
2. *Data Selection*
   Tahapan *Selection* digunakan untuk mengidentifikasi variabel yang akan diambil, dengan tujuan menghindari duplikasi dan mengurangi perulangan yang tidak diperlukan dalam proses pengolahan data.
3. *Data Preprocessing*
   Tahap *Preprocessing* digunakan untuk menjalankan proses pembersihan atau *cleaning* pada data yang digunakan dalam konteks penelitian.
4. *Data Transformation*
   Transformasi data dilakukan untuk menyesuaikan format data dengan persyaratan pengolahan dalam *data mining*. Hal ini diperlukan karena beberapa metode dalam *data*

mining memerlukan struktur data tertentu sebelum dapat dijalankan secara efektif.

5. *Data Mining*
   Proses inti dalam metode yang diterapkan untuk memperoleh informasi baru dari data yang telah diolah. *Data mining* merupakan suatu teknik yang digunakan untuk mengidentifikasi pola atau informasi yang signifikan dalam suatu himpunan data. Tujuan dan penemuan pengetahuan dalam proses *Knowledge Discovery in Databases (KDD)* menetapkan metode, teknik, dan algoritma penambangan data yang optimal. Dalam penelitian ini, proses *data mining* yang diterapkan adalah *Clustering* menggunakan Algoritma *K-Means.*
6. *Evaluation*
   Pada tahap ini, metrik evaluasi yang digunakan dalam penelitian ini adalah *Davies-Bouldin Index* (DBI). Penggunaan *Davies-Bouldin Index* (DBI) untuk menentukan jumlah *cluster* terbaik. Nilai DBI yang terkecil atau paling rendah menunjukkan jumlah *cluster* yang lebih optimal.
7. *Knowledge*
   Pola-pola yang dihasilkan akan disajikan kepada pengguna untuk menjadi pengetahuan lebih dalam.

4. HASIL DAN PEMBAHASAN

Hasil penelitian yang dilakukan dalam pembahasan ini akan menjelaskan bagaimana proses *clusterisasi* atau mengelompokkan *dataset* lagu-lagu paling populer di *Spotify* 2023 berdasarkan suasana hati dengan menggunakan Algoritma *K-Means.* Penelitian ini dilaksanakan dengan proses pengujian *machine learning* dengan *Software RapidMiner* versi 10.2.

4.1. Data

Data yang digunakan dalam bentuk *file excel* dengan judul *Most Streamed Spotify Songs* 2023 diperoleh dari *Website Kaggle*. Jumlah *record* data sebanyak 952 *record*, terdiri dari 24 atribut. Untuk memasukkan *dataset* ke dalam *repository RapidMiner*, langkahnya adalah *mengimport* data *Most Streamed Spotify Songs* 2023.xlsx ke dalam operator *Retrieve*, sesuai dengan tampilan yang terlihat pada Gambar 3 di bawah ini.

Gambar 3. Operator *Retrieve* pada *RapidMiner*

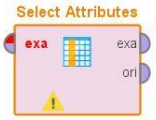
Hasil dari penggunaan operator *Retrive* pada *RapidMiner* terdapat pada Tabel 1 yang diperoleh informasi :

Tabel 1. Hasil dari Operator *Retrieve*

No.	Uraian	Keterangan
1.	*Record*	952
2.	*Special Attributes*	0
3.	*Reguler Attributes*	24

4.2. *Data Selection*

Operator di *RapidMiner* yang digunakan untuk seleksi data adalah operator *Select Attributes* dan operator *Set Role*. Operator *Select Attributes* pada Gambar 4 berfungsi untuk menentukan atribut yang akan digunakan dalam proses analisis atau pemodelan.



Gambar 4. Operator *Select Atributtes*

Operator ini memilih *subset* Atribut dari *ExampleSet* dan menghapus atribut yang tidak dipilih. Operator menyediakan berbagai jenis filter untuk menyederhanakan proses pemilihan atribut. Parameter pada operator *Select Attributes* yang digunakan terdapat pada Tabel 2 di bawah ini.

Tabel 2. *Parameter* Operator *Select Attributes*

No.	*Parameter*	Isi
1.	*type*	*include attributes*
2.	*attribute filter type*	*a subset*
3.	*select subset*	*select attributes :*
		bpm
		danceability_%
		energi_%
		track_name
		valence_%

Penjelasan dari masing-masing atribut yang di gunakan terdapat pada Tabel 3 yang diperoleh dari Sumber data [13].

Tabel 3. Atribut yang digunakan

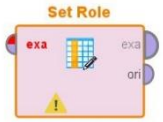
Atribut	Keterangan
track_name	Judul lagu
Bpm	Denyut per menit, ukuran tempo lagu
Danceability	Seberapa cocok lagu tersebut untuk menari
Valence	Kepositifan konten musik lagu
Energy	Tingkat energi yang dirasakan dari lagu tersebut

Kisaran nilai setiap atribut dikatakan sebagai tinggi, sedang dan rendah dan suasana hati yang mewakili ada pada Tabel 4 di bawah ini :

Tabel 4. Keterangan Suasana Hati

Bpm	*Danceability*	*Valence*	*Energy*	Ket.	Suasana Hati
>150	>70	>75	>80	Tinggi	*Happy*
90-150	40-70	50-75	40-80	Sedang	*Relax*
< 90	< 40	< 50	< 40	Rendah	*Sad*

Selanjutnya untuk menentukan *id dataset* menggunakan operator *Set Role* terdapat pada Gambar 5 di bawah ini.



Gambar 5. Operator *Set Role*

Parameter yang digunakan pada operator *Set Role* tampak pada Tabel 5 di bawah ini :

Tabel 5. Parameter Operator *Set Role*

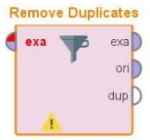
No.	Parameter	Isi
1.	*attribute name*	*track_name*
2.	*target role*	*id*

4.3. *Data Preprocessing*

Proses *Preprocessing* merupakan langkah yang diterapkan untuk melakukan pembersihan pada data yang digunakan dalam konteks penelitian. Tahapan ini bertujuan untuk menghapus atribut data yang mengandung duplikat, pada *RapidMiner* operator yang digunakan untuk proses ini adalah operator *Remove Duplicates*. Berdasarkan Gambar 6 yang merupakan hasil *result* dari statistik *dataset* yang mengandung nilai duplikat.

Gambar 6. *Result* dari statistik *dataset*

Berdasarkan Gambar 6 di atas mengandung nilai duplikat karena data mentah yang ada terdiri dari 952 *record* tampak pada Gambar 4.6, sedangkan pada *result* statistik datanya hanya 942 *record*. Untuk mengatasi nilai yang mengandung duplikat pada *RapidMiner* menggunakan operator *Remove Duplicates* terdapat pada Gambar 7 di bawah ini.



Gambar 7. Operator *Remove Duplicates*

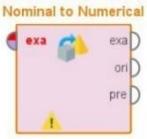
Parameter yang digunakan pada operator *Remove Duplicates* terdapat pada Tabel 6 sebagai berikut :

Tabel 6. Parameter Operator *Remove Duplicates*

No.	Parameter	Isi
1.	*attribute filter type*	*subset*
2.	*attributes*	*Select Attributes :*
		track_name
3.	*include special attributes*	centang

4.4. *Data Transformation*

Karena tipe data yang tersedia terdapat tipe data *non-numeric* sedangkan Algoritma *K-Means Clustering* harus bertipe *numeric*, maka diperlukan *transformation* untuk mengubah data bertipe *polynominal* menjadi *numeric* menggunakan operator *Nominal to Numeric* di *Rapidminer* seperti pada Gambar 8 di bawah ini.



Gambar 8. Operator *Nominal to Numeric*

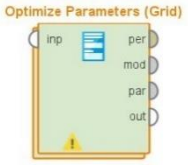
Parameter pada operator *Nominal to Numeric* yang digunakan ada pada Tabel 7 di bawah ini.

Tabel 7. Parameter Operator *Nominal to Numeric*

No.	Parameter	Isi
1.	*attribute filter type*	*subset*
2.	*attributes*	*Select Attributes :*
		track_name

4.5. *Data Mining*

Dalam penelitian ini, proses *data mining* yang diterapkan adalah *Clustering* menggunakan Algoritma *K-Means* , yang diaplikasikan untuk mengelompokkan data lagu terpopuler di *Spotify* tahun 2023. Sebelum menerapkan operator *K-Means* di *RapidMiner*, langkah pertama adalah menggunakan operator *Optimize Parameter (Grid*) tampak pada Gambar 9 di bawah. Hal ini dilakukan untuk mencari kombinasi parameter *K-Means* yang optimal guna meningkatkan kinerja klasterisasi.



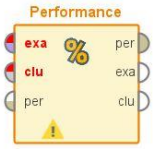
Gambar 9. Operator *Optimize Parameter (Grid)*

Setelah menerapkan Operator *optimize Parameter (Grid)*, masukan operator *K-Means* yang tampak pada Gambar 10 digunakan untuk menerapkan algoritma dengan parameter terbaik yang ditemukan.

Gambar 10. Operator *K-Means*

Selanjutnya, tambahkan Operator *Cluster Distance Performance* yang terdapat pada Gambar 11 untuk mengevaluasi kualitas *cluster* dengan metrik evaluasi tertentu, memberikan wawasan tentang seberapa baik model *K-Means* menyesuaikan diri

dengan data dan sejauh mana *cluster-cluster* tersebut saling terpisah.



Gambar 11. Operator *Cluster Distance Performance*

Parameter operator *K-Means* dan operator *Cluster Distance Performance* yang digunakan dan diterapkan pada Operator *Optimize Parameter (Grid)* terdapat pada Tabel 8 di bawah ini.

Tabel 8. Parameter yang digunakan

No.	Operator	Parameter	*Select Parameter*
1.	*Clustering (k-Means)*	*Clustering.k*	*Min = 2.0*
			Max = 10.0
			Steps = 100
			Scale = linear
		Clustering.measure_type	*MixedMeasures*
			NumericalMeasures
			BregmanDivergences
2.	*Cluster Distance Performance*	*Performance.main_criterion*	*Davies Bouldin*

4.6. *Evaluation*

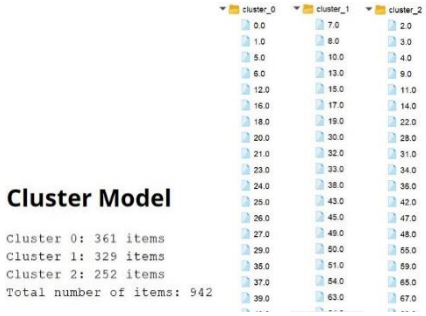
Davies-Bouldin Index (DBI) digunakan untuk menentukan jumlah kluster terbaik. Nilai DBI yang terkecil atau paling rendah menunjukkan jumlah *cluster* yang lebih optimal. Dalam penelitian ini klaster terbaik ada di k = 3 dengan nilai DBI sebesar 0,299 terdapat pada Tabel 9 di bawah ini.

Tabel 9. Komparasi nilai DBI

iteration	*Clustering.k*	*Clustering.measure_type*	*Performance.main_criterion*	*DBI*
2	3	*BregmanDivergences*	*Davies Bouldin*	-0.299
11	3	*NumericalMeasures*	*Davies Bouldin*	-0.299
12	4	*NumericalMeasures*	*Davies Bouldin*	-0.305
14	6	*NumericalMeasures*	*Davies Bouldin*	-0.310
20	3	*MixedMeasures*	*Davies Bouldin*	-0.299
21	4	*MixedMeasures*	*Davies Bouldin*	-0.305
24	7	*MixedMeasures*	*Davies Bouldin*	-0.309
26	9	*MixedMeasures*	*Davies Bouldin*	-0.309

4.7. *Knowledge*

Selanjutnya dari proses *K-Means Clustering* diperoleh *Cluster* Model dan anggota setiap *cluster* terdapat pada Gambar 12 di bawah ini.



Gambar 12. *Cluster* model dan anggotanya

Dari Gambar 12 dapat dilihat hasil dari *Cluster* Model dan anggota setiap *cluster* dapat dianalisis bahwa distribusi elemen-elemen dalam kluster menunjukkan bahwa data telah berhasil dikelompokkan menjadi tiga kluster (*Cluster* 0 : 361 anggota, *Cluster* 1 : 329 anggota, dan *Cluster* 2 : 252 anggota) dengan total 942 data. Jumlah item yang berbeda di setiap *cluster* menunjukkan adanya variasi dalam sifat atau karakteristik item-item tersebut, seperti kesamaan atau pola tertentu yang ditemukan oleh Algoritma *K-Means Clustering*. *Cluster* dengan jumlah item yang lebih tinggi menunjukkan adanya lebih banyak kesamaan di antara elemen-elemen tersebut, sedangkan kluster dengan jumlah item yang lebih rendah menunjukkan variasi yang lebih tinggi. Analisis lebih lanjut diperlukan untuk memahami makna atau signifikan *cluster-cluster* ini dalam konteks tujuan analisis

Selanjutnya terdapat tabel *centroid* yang di dapat dari proses *K-Means Clustering* di *tools RapidMiner* tampak pada Gambar 13.

Attribute	cluster_0	cluster_1	cluster_2
bpm	116.673	101.657	157.544
danceability_%	75.859	63.271	59.048
valence_%	72.249	33.903	44.393
energy_%	72.064	57.954	61.627

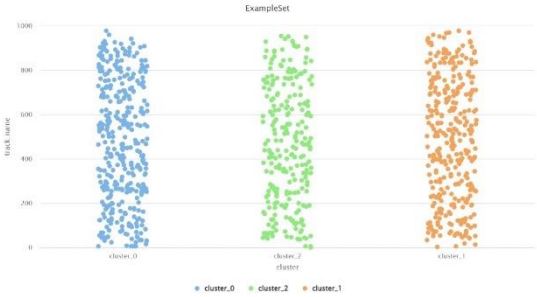
Gambar 13. Tabel centroid

Pada Gambar 13 dapat dianalisis bahwa hasil dari Tabel *Centroid* dapat dianalisis sebagai berikut :

1. *Cluster* 0 dengan BPM yang sedang, *Danceability* yang tinggi, *Valence* yang tinggi, dan Energy tinggi dapat diinterpretasikan sebagai suasana hati "*Happy*" atau "Senang"

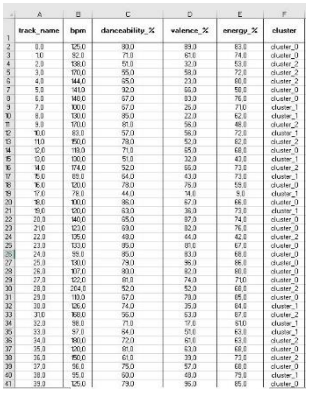
2. *Cluster* 1 dengan BPM yang rendah, *Danceability* yang sedang, *Valence* yang rendah, dan Energi yang rendah dapat diinterpretasikan sebagai suasana hati “Sad” atau “Sedih”.
3. *Cluster* 2 dengan BPM yang tinggi, *Danceability* yang rendah, *Valence* yang sedang, dan Energi yang sedang dapat diinterpretasikan sebagai suasana hati “*Relax*” atau “Santai”.
4. Dari tabel centroid yang diberikan, tidak ada *cluster* yang secara jelas dapat diinterpretasikan sebagai suasana hati “marah”.

Visualisasi dari hasil k-means *cluster*ing antara *cluster* dan atribut track_name tampak pada Gambar 14 di bawah ini：



Gambar 14. Visualisasi *cluster* dan *track_name*

Berdasarkan Visualisasi *cluster* dan atribut *track_name* dapat di analisis bahwa anggota *Cluster* 2 memiliki distribusi yang lebih baik dibandingkan *cluster* 0 dan *cluster* 1, distribusi anggota di setiap *cluster* memiliki jumlah yang mirip, tanpa menonjolkan perbedaan yang mencolok. Dan peringkat *cluster* 2 lebih baik dibandingkan *cluster* 0 dan *cluster* 1. Hasil pemetaan *cluster* yang telah diubah ke dalam format *excel* (xlsx) akan ditampilkan sebagaimana dilihat pada Gambar 15.



	A	B	C	D	E	F
1	track_name	bpm	danceability_%	valence_%	energy_%	cluster
2	0.0	125.0	90.0	89.0	83.0	cluster_0
3	1.0	92.0	71.0	61.0	74.0	cluster_0
4	2.0	138.0	51.0	32.0	53.0	cluster_2
5	3.0	170.0	55.0	58.0	72.0	cluster_2
6	4.0	144.0	65.0	23.0	80.0	cluster_2
7	5.0	141.0	32.0	66.0	58.0	cluster_0
8	6.0	148.0	67.0	83.0	76.0	cluster_0
9	7.0	100.0	67.0	26.0	71.0	cluster_1
10	8.0	130.0	65.0	22.0	62.0	cluster_1
11	9.0	170.0	67.0	56.0	68.0	cluster_2
12	10.0	83.0	57.0	56.0	72.0	cluster_1
13	11.0	150.0	79.0	52.0	62.0	cluster_2
14	12.0	118.0	71.0	95.0	68.0	cluster_0
15	13.0	120.0	51.0	32.0	40.0	cluster_1
16	14.0	174.0	52.0	66.0	73.0	cluster_2
17	15.0	89.0	54.0	43.0	72.0	cluster_1
18	16.0	120.0	78.0	76.0	59.0	cluster_0
19	17.0	76.0	44.0	14.0	9.0	cluster_1
20	18.0	100.0	86.0	67.0	66.0	cluster_0
21	19.0	130.0	63.0	36.0	73.0	cluster_1
22	20.0	140.0	95.0	97.0	74.0	cluster_0
23	21.0	123.0	69.0	82.0	76.0	cluster_0
24	22.0	135.0	69.0	44.0	42.0	cluster_2
25	23.0	132.0	95.0	81.0	67.0	cluster_0
26	24.0	98.0	85.0	83.0	68.0	cluster_0
27	25.0	130.0	79.0	96.0	66.0	cluster_0
28	26.0	107.0	80.0	82.0	66.0	cluster_0
29	27.0	122.0	61.0	74.0	71.0	cluster_0
30	28.0	154.0	52.0	52.0	68.0	cluster_2
31	29.0	110.0	67.0	79.0	65.0	cluster_0
32	30.0	126.0	74.0	35.0	64.0	cluster_1
33	31.0	168.0	56.0	53.0	67.0	cluster_2
34	32.0	98.0	71.0	17.0	51.0	cluster_1
35	33.0	97.0	64.0	25.0	63.0	cluster_1
36	34.0	180.0	72.0	61.0	63.0	cluster_2
37	35.0	120.0	61.0	53.0	68.0	cluster_0
38	36.0	150.0	61.0	33.0	72.0	cluster_2
39	37.0	96.0	75.0	57.0	68.0	cluster_0
40	38.0	95.0	60.0	43.0	73.0	cluster_1
41	39.0	125.0	79.0	96.0	85.0	cluster_0

Gambar 15. Hasil data *export* pemetaan *cluster*

Pada Gambar 15 dapat dilihat hasil dari data *export* pemetaan *cluster* bahwa penelitian ini telah berhasil mengelompokkan lagu-lagu terpopuler Spotify dengan *K-Means.*

Analisis klasterisasi menunjukkan bahwa lagu-lagu terpopuler dibagi ke dalam kluster berdasarkan variasi suasana hati, seperti "*Happy*," "Sad," dan "*Relax*." Hal ini menandakan bahwa pendengar dapat menemukan variasi musik yang sesuai dengan berbagai suasana hati atau emosi, menciptakan pengalaman mendengarkan yang lebih beragam dan kaya secara emosional.

Tabel centroid memberikan interpretasi tentang bagaimana atribut-atribut seperti *Bpm*, *Danceability*, *Valence*, dan Energy berkaitan dengan suasana hati. Misalnya, *cluster* 0 dengan *Bpm* yang sedang, *Danceability* yang tinggi, *Valence* yang tinggi, dan Energi yang tinggi dapat diinterpretasikan sebagai suasana hati "*Happy*". Ini menunjukkan bahwa karakteristik musik dapat memengaruhi cara kita merasakan dan mengalami suasana hati yang terkait.

Melalui Tabel yang memberikan pengertian tentang atribut audio seperti *Bpm*, *Danceability*, *Valence*, dan Energy, kita dapat mengamati bagaimana nilai-nilai tertentu dari atribut tersebut berkorelasi dengan suasana hati yang dihasilkan. Ini memberikan wawasan lebih lanjut tentang bagaimana elemen-elemen musik dapat memengaruhi pengalaman mendengarkan dan respons psikologis-emosional pendengar.

Dengan demikian, hasil klasterisasi ini memberikan bukti bahwa variasi atribut musik yang terdapat dalam kluster memengaruhi pengalaman mendengarkan musik dan memiliki hubungan erat dengan aspek psikologis dan emosional dalam konsep musik digital. Distribusi yang seimbang, interpretasi atribut, dan visualisasi kluster semuanya berkontribusi untuk mendukung hubungan ini dan memberikan dasar yang kuat bagi pengaruh musik terhadap pengalaman pendengar.

Berdasarkan jumlah anggota di setiap *cluster* dapat disimpulkan bahwa hasil *cluster*ing menunjukkan distribusi anggota yang relatif seimbang di antara tiga *cluster*. Jumlah anggota yang mendekati nilai yang serupa pada setiap *cluster* (*cluster* 0 : 361, *cluster* 1 : 329, dan *cluster* 2 : 252) menunjukkan bahwa tidak ada satu *cluster* pun yang mendominasi secara signifikan dalam hal jumlah anggota. Dengan kata lain, distribusi anggota di antara *cluster* tidak terlalu jauh, dan tidak ada *cluster* yang secara drastis lebih besar atau lebih kecil dari yang lain.

Distribusi anggota yang relatif seimbang di antara *cluster* menunjukkan bahwa tidak ada satu jenis suasana hati yang mendominasi secara signifikan. Pendengar memiliki akses ke berbagai jenis musik yang mencerminkan berbagai aspek emosional dan psikologis. Ini memberikan pilihan yang lebih luas dan mendalam dalam menciptakan pengalaman mendengarkan yang sesuai dengan keadaan emosional dan psikologis individu.

5. KESIMPULAN DAN SARAN

Kesimpulan pada penelitian ini dengan menggunakan *K-Means Clustering* dengan k=3 pada lagu-lagu *Spotify* 2023 menunjukkan klasterisasi yang baik dengan *Davies-Bouldin Index* (DBI) 0,299 pada iterasi kedua. Analisis suasana hati menggunakan

model *Thayer's* mengidentifikasi tiga *cluster* utama: *Cluster* 0 ("*Happy*") dengan *Bpm* sedang, *Danceability* tinggi, *Valence* tinggi, dan *Energy* tinggi; *Cluster* 1 ("*Sad*") dengan *Bpm* rendah, *Danceability* sedang, *Valence* rendah, dan *Energy* rendah; *Cluster* 2 ("*Relax"*) dengan *Bpm* tinggi, *Danceability* rendah, *Valence* sedang, dan *Energy* sedang. Tidak ada *cluster* yang khusus menggambarkan "*Angry*". Distribusi anggota yang merata menunjukkan variasi suasana hati. Saran dalam penelitian selanjutnya termasuk penambahan atribut audio, validasi dengan metode lain, melibatkan ahli musik, dan mengulangi analisis pada *dataset* berbeda untuk pemahaman yang lebih komprehensif.

DAFTAR PUSTAKA

[1] S. Navisa, L. Hakim, and A. Nabilah, "Komparasi Algoritma Klasifikasi Genre Musik pada Spotify Menggunakan CRISP-DM," vol. 04, no. 02, pp. 114–125, 2021.

[2] P. I. Maulana, A. Aranta, F. Bimantoro, and I. G. Andika, "KLASIFIKASI MOOD MUSIK BERDASARKAN MEL FREQUENCY CEPSTRAL," vol. 4, no. 1, pp. 72–85, 2022.

[3] F. F. Surenggana, A. Aranta, and F. Bimantoro, "KLASIFIKASI MOOD MUSIK MENGGUNAKAN K-NEAREST NEIGHBOR DENGAN MEL FREQUENCY CEPSTRAL COEFFICIENTS (Mood Music Classification using K-Nearest Neighbor with Mel Frequency Cepstral," vol. 4, no. 2, pp. 263–276, 2022.

[4] E. T. Andaryani, "PENGARUH MUSIK DALAM MENINGKATKAN MOOD BOOSTER MAHASISWA THE EFFECTS OF MUSIC IN IMPROVING STUDENT ' S MOOD BOOSTER Pendahuluan Pembahasan," vol. 1, pp. 109–115, 2019.

[5] N. A. Saputri, "Ekonomi politik media dalam industri musik digital spotify," pp. 6–11, 2021.

[6] Rubangiya, T. Hartati, and Y. A. Wijaya, "ANALISIS DATA LALU LINTAS JARINGAN DI KANTOR CANGEHGAR CYBER OPERATION CENTER MENGGUNAKAN ALGORITMA K-MEANS NETWORK TRAFFIC DATA ANALYSIS AT CANGEHGAR CYBER OPERATION CENTER OFFICE USING K-MEANS ALGORITHM," vol. 7, no. 1, pp. 75–84, 2022.

[7] L. Nurhalimah, T. I. Hermanto, and I. Kaniawulan, "Analisis Prediksi Mood Genre Musik Pop Menggunakan Algoritma," vol. 9, no. 4, pp. 1006–1013, 2022, doi: 10.30865/jurikom.v9i4.4597.

[8] Gustientiedina, M. H. Adiya, and Y. Desnelita, "Jurnal Nasional Teknologi dan Sistem Informasi Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru," vol. 01, pp. 17–24, 2019.

[9] S. Handoko, Fauziah, and E. T. E. Handayani, "IMPLEMENTASI DATA MINING UNTUK MENENTUKAN TINGKAT PENJUALAN PAKET DATA TELKOMSEL MENGGUNAKAN METODE K-MEANS CLUSTERING," vol. 25, no. 1, 2020.

[10] H. Kurniawan, S. Defit, and Sumijan, "JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST) Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal," vol. 1, no. 2, pp. 80–89, 2020.

[11] R. Shr, "8 Pengertian Mood Menurut Pendapat Ahli, Beda dengan Emosi," *IDN TIMES*, 2023. https://www.idntimes.com/science/discovery/amp/indriani-s-1/perbedaan-mood-dengan-emosi-c1c2?page=all#page-2 (accessed Nov. 25, 2023).

[12] N. I. Febianto and N. D. Palasara, "Analisis Clustering K-Means Pada Data Informasi Kemiskinan Di Jawa Barat Tahun 2018," vol. 08, no. September, pp. 130–140, 2019.

[13] N. Elgiriyewithana, "Most Streamed Spotify Songs 2023," *Kaggle*, 2023. https://www.kaggle.com/datasets/nelgiriyewithana/top-spotify-songs-2023 (accessed Nov. 01, 2023).