

## Penerapan Algoritma Klustering K-Means Untuk Fitur Atribut Pada Layanan Streaming Musik Spotify

Muhammad Ikhsan Firmansyah<sup>1</sup>, Ramdhan Saepul Rohman<sup>2</sup>,  
Eva Marsusanti<sup>3</sup>

Universitas Bina Sarana Informatika <sup>1,2,3</sup>

Ikhsanfirmsyah24@gmail.com<sup>1</sup>, ramdhan.rpe@bsi.ac.id<sup>2</sup>, eva.emr@bsi.ac.id<sup>3</sup>

---

Diterima (22-09-2023)	Direvisi (29-09-2023)	Disetujui (10-10-2023)
--------------------------	--------------------------	---------------------------

---

**Abstrak** - Penelitian ini mencoba mencari fitur atribut pada track lagu yang memiliki jumlah paling banyak pada setiap lagu pada aplikasi spotify, sekaligus mencoba menerapkan metode kluster pada dataset yang diperoleh dari kaggle sebagai data pabrik. Spotify merupakan salah satu aplikasi streaming musik yang paling diminati oleh beberapa pendengar. Menggunakan algoritma k-means dan *davies Bouldin Indeks*(DBI) sebagai metode validasi dari hasil klustering k-means. Adapun metode yang digunakan dalam penelitian ini adalah CRISP-DM sebagai standarisasi pengolahan data mining lintas industri yang dinilai cocok dalam melakukan penelitian dan proses data mining, dari melakukan pemahaman bisnis sampai melakukan tahap evaluasi. Hasil didapatkan dari penelitian berupa atribut yang berpengaruh pada lagu adalah *instrumentalness* dan *valance* terhadap lagu yang populer di aplikasi spotify serta atribut yang secara keseluruhan seimbang dari antar kluster adalah *loudness*. Dengan dataset lagu yang popularitas di atas 60, genre pop menjadi yang paling banyak dalam jumlah deretan lagunya. Pada percobaan kluster dengan menggunakan algoritma k-means serta evaluasi *davies Bouldin indeks*(DBI) memperoleh jumlah kluster yang dibagi menjadi 9 adalah yang paling optimal dari percobaan pembagian jumlah kluster dari 2 kelompok sampai 10 kelompok. Pada penelitian ini akan dilakukan dengan menggunakan aplikasi rapidminer sebagai alat bantu peneliti untuk melakukan pengujian dan perhitungan.

Kata Kunci : Data mining, Spotify, Lagu, K-means, Davies Bouldin Index

**Abstract** - This research attempts to find attribute features in the track songs that have the highest count for each song on the Spotify application, while also attempting to apply clustering methods to the dataset obtained from Kaggle as factory data. Spotify is one of the most popular music streaming applications among listeners. The research uses the K-means algorithm and Davies Bouldin Index (DBI) as the validation method for the K-means clustering results. The method used in this study is CRISP-DM, which is a standardization of cross-industry data mining processing that is considered suitable for conducting research and data mining processes, from business understanding to evaluation stages. The results obtained from the research are attributes that influence the songs, such as *instrumentalness*, and an overall balanced attribute among the clusters, which is *loudness*. With a dataset of songs with a popularity above 65, the pop genre is the most abundant in terms of the number of songs. In the clustering experiment using the K-means algorithm and evaluating the Davies Bouldin Index (DBI), the optimal number of clusters obtained is 9, which is divided into 2 to 10 clusters. This research will be conducted using the RapidMiner application as a researcher's tool for testing and calculations.

Keywords: Data mining, Spotify, Songs, K-means, Davies Bouldin Index

### I. PENDAHULUAN

Sebelumnya, menikmati musik memerlukan proses yang rumit. Namun, saat ini, musik dapat dinikmati dengan mudah melalui berbagai platform digital. Musik juga dapat digunakan sebagai sarana hiburan dan konektivitas sosial (Dewatara & Agustin, 2019) Aplikasi *streaming* musik telah menjadi sarana utama untuk menikmati musik di banyak bagian dunia. *Streaming* berdampak pada artis, ekonomi industri musik, dan musik itu sendiri. Ada spekulasi tentang dampak pergeseran ini

pada distribusi kekuasaan. *Streaming* dapat membebaskan distribusi dan penemuan musik dari label besar dan memberi lebih banyak ruang bagi artis independen (Prey et al., 2022) Selain Spotify, Apple Music juga merupakan layanan *streaming* musik populer. Diluncurkan pada tahun 2015, Apple Music telah memiliki lebih dari 40 juta pelanggan berbayar di seluruh dunia. Pasar utamanya adalah Amerika Serikat, tetapi juga kompetitif di Eropa. Apple Music mengenakan harga yang sama dengan Spotify untuk langganan bulanan (Vonderau, 2019)

Faktor penting dalam bisnis *streaming* musik adalah rekomendasi musik yang akurat. Rekomendasi yang akurat akan meningkatkan kepuasan pengguna, meningkatkan pendapatan, dan menarik lebih banyak pengguna. Spotify dan Apple Music terus menyempurnakan sistem rekomendasi mereka untuk memberikan pengalaman yang lebih baik bagi pengguna (Florez Ramos & Blind, 2020).

Dengan mudahnya mendengarkan musik ini memberikan gaya hidup baru kepada masyarakat, dapat diakses melalui *smartphone* maupun Laptop/PC secara fleksibel. Dan kumpulan musik di spotify ini dikumpulkan dan dikelola dalam sebuah *database* milik penyedia layanan dan diproses pemeliharaan dan pengurutan lagu yang sesuai dengan lagu yang sering kita dengarkan berdasarkan nama penyanyi, album, genre lagu dan hal lainnya.

Setiap lagu memiliki karakteristik nada yang unik, seperti *danceability, energy, loudness, speechiness, acousticness, instrumentality, liveness, valence, dan tempo*.

Metode CRISP-DM penting digunakan dalam penelitian ilmiah dengan data mining karena dapat membantu peneliti untuk melakukan data mining secara sistematis dan terstruktur. Kerangka kerja ini dapat membantu peneliti untuk menghindari kesalahan dalam proses data mining, sehingga hasil penelitian menjadi lebih akurat dan valid. Algoritma k-means adalah salah satu algoritma clustering yang populer dalam data mining. Algoritma ini digunakan untuk mengelompokkan data menjadi beberapa kelompok berdasarkan kesamaannya. Algoritma k-means penting digunakan dalam penelitian ilmiah dengan data mining karena dapat membantu peneliti untuk menemukan pola dan kecenderungan dalam data. Pola dan kecenderungan ini dapat digunakan untuk menarik kesimpulan dan membuat rekomendasi. Serta dibantu menggunakan evaluasi dari DBI (*Davies Bouldin Indeks*) dalam menemukan jumlah kluster yang tepat dalam penelitian ini.

### 1. Tujuan Penelitian

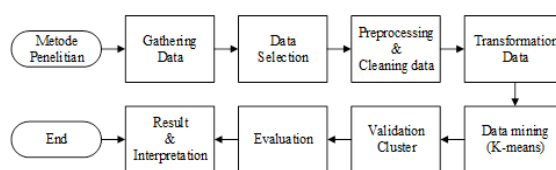
Penelitian ini memiliki tujuan seperti berikut:

1. Melakukan klustering data menggunakan algoritma K-Means untuk mengetahui fitur audio atribut yang selalu muncul di aplikasi spotify.
2. Penerapan klustering pada aplikasi rapidminer menggunakan K-means pada hasil jumlah kluster beserta anggota kluster.
3. Menggunakan evaluasi DBI(Davies Bouldin Indeks) untuk menentukan jumlah kluster yang paling optimal dari beberapa percobaan.

4. Menemukan genre dengan yang memiliki jumlah lagu terbanyak.

### 2. Gambaran Metode Usulan

Dari beberapa penelitian terdahulu, dalam melakukan klustering dalam proses penelitiannya. Rata-rata memiliki urutan penelitian seperti pengumpulan data, implementasi, uji dan analisis serta simpulan. Tidak melakukan evaluasi sebelum melakukan deployment data. Seperti melakukan evaluasi dalam menentukan jumlah kluster dalam penelitiannya. Dengan menggunakan Davies Bouldin Index maka dapat mengetahui pengolahan data dengan jumlah kluster berapa yang cocok dilakukan dalam penelitian.



Sumber: Alur Penelitian (2023)

Gambar 1. Alur Metode Penelitian

1. Pengumpulan data: Tahap ini melibatkan pencarian dan pengumpulan data dari berbagai sumber, baik internal maupun eksternal. Data dapat berasal dari berbagai format, seperti file teks, database, atau media sosial.
2. Pemilihan data: Tahap ini melibatkan pemilihan data yang relevan untuk dianalisis. Data yang dipilih harus sesuai dengan tujuan dari proses data mining.
3. Pembersihan data: Tahap ini melibatkan perbaikan kesalahan dan tidak konsisten data. Data yang tidak akurat atau tidak konsisten dapat menyebabkan hasil analisis yang tidak valid.
4. Transformasi data: Tahap ini melibatkan mengubah format data agar sesuai untuk analisis. Transformasi ini sering melibatkan peringkasan atau penyederhanaan data.
5. Penambangan data: Tahap ini melibatkan pencarian pola dan informasi dalam data. Pola dan informasi ini dapat digunakan untuk membuat keputusan atau meningkatkan kinerja.
6. Evaluasi data: Tahap ini melibatkan pemeriksaan kebenaran dan kegunaan pola yang ditemukan. Pola yang ditemukan harus benar dan dapat digunakan untuk mencapai tujuan dari proses data mining.

### II. METODOLOGI PENELITIAN

Penelitian kuantitatif menggunakan data berupa angka untuk menggambarkan karakteristik objek, peristiwa, atau situasi. Pendekatan ini didasarkan pada filsafat positivisme dan sering

digunakan untuk menguji hipotesis (Mukhid, 2021).

Data akan diolah oleh proses data mining, Data mining digunakan untuk mengungkapkan informasi penting yang tersembunyi dari *database* atau sumber data yang besar. Kemajuan dalam berbagai bidang ilmu seperti sains, bisnis, dan lainnya telah menghasilkan koleksi *database* yang terus berkembang. Kumpulan data yang besar ini dapat dimanfaatkan untuk mendukung proses pengambilan keputusan (Jollyta, n.d., 2020).

Pengumpulan data yang digunakan adalah data sekunder, Sumber informasi sekunder dalam penelitian ini adalah buku, jurnal, dan artikel tentang sistem pengendalian internal sistem penggajian (Uswatun, 2021)

terutama untuk mencari dataset yang akan diolah. Tidak berdasarkan dataset secara langsung yang dimiliki instansi khusus, sehingga pengumpulan data ini menggunakan data sekunder.

Sumber data ini berasal dari forum kaggle, forum kaggle ini berisi semua dataset public Kaggle adalah platform yang menyelenggarakan kompetisi data sains untuk masalah bisnis, rekrutmen, dan tujuan penelitian akademis. Berikut adalah definisi dari setiap atribut dataset yang akan digunakan (variabel), Data yang akan digunakan adalah data yang dimiliki dari Maharshi Pandya sebagai author dan collaborators (kaggle.com):

Tabel 1. Definisi dataset penelitian

Atribut	Definisi
<i>popularity</i>	Popularitas lagu dihitung berdasarkan jumlah pemutaran dan seberapa baru pemutaran tersebut. Lagu yang sering diputar saat ini akan lebih populer daripada lagu yang sering diputar di masa lalu. Track duplikat akan dinilai secara terpisah. Popularitas artis dan album juga dihitung berdasarkan popularitas track yang terkait.
<i>danceability</i>	<i>Danceability</i> adalah ukuran seberapa cocok sebuah lagu untuk ditari, berdasarkan tempo, ritme, ketukan, dan keteraturan. Nilai <i>danceability</i> berkisar antara 0,0 hingga 1,0.
<i>energy</i>	<i>Energy</i> adalah ukuran intensitas dan aktivitas sebuah lagu. Lagu dengan energi tinggi terasa cepat, keras, dan berisik. Contoh lagu dengan energi tinggi adalah death metal, sedangkan prelude Bach memiliki energi yang rendah. berkisar antara 0,0

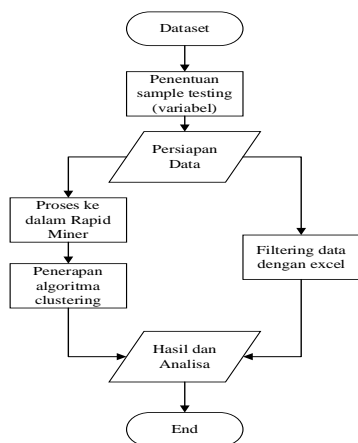
<i>loudness</i>	hingga 1,0. Kenyaringan keseluruhan track dalam desibel (dB)
<i>speechiness</i>	<i>Speechiness</i> adalah ukuran seberapa banyak kata yang diucapkan dalam sebuah lagu. Nilainya berkisar antara 0,0 hingga 1,0. Lagu dengan <i>speechiness</i> tinggi kemungkinan besar terdiri dari kata-kata yang diucapkan, sedangkan lagu dengan <i>speechiness</i> rendah kemungkinan besar terdiri dari musik.
<i>acousticness</i>	<i>Confidence</i> digunakan untuk mengukur apakah sebuah lagu adalah akustik, dengan rentang nilai antara 0,0 hingga 1,0. Nilai 1.0 menunjukkan keyakinan yang tinggi bahwa lagu tersebut memang merupakan lagu akustik.
<i>instrumentalness</i>	Prediksi dilakukan untuk menentukan apakah sebuah lagu tidak mengandung vokal. Bunyi "Ooh" dan "aah" dianggap instrumen, sedangkan lagu rap dan lagu dengan kata-kata diucapkan dianggap vokal. Nilai instrumental yang mendekati 1.0 menunjukkan lagu tidak mengandung vokal.
<i>liveness</i>	Nilai keaktifan penonton dapat digunakan untuk mendeteksi apakah rekaman direkam secara langsung. Nilai di atas 0,8 menunjukkan kemungkinan yang kuat bahwa rekaman tersebut diambil selama konser berlangsung.
<i>valence</i>	<i>Valence</i> mengukur tingkat kepositifan sebuah lagu. Lagu dengan <i>valence</i> tinggi positif, sedangkan lagu dengan <i>valence</i> rendah negatif berkisar antara 0,0 hingga 1,0.
<i>tempo</i>	Estimasi tempo lagu mengacu pada kecepatan secara keseluruhan dalam ketukan per menit (BPM).
<i>track_genre</i>	Genre yang dimiliki oleh lagu

Sumber: Kaggle.com

Algoritma K-means adalah salah satu algoritma clustering yang populer dalam data mining. Algoritma ini digunakan untuk mengelompokkan data menjadi beberapa kelompok berdasarkan kesamaannya. Metode CRISP-DM adalah kerangka kerja yang terstruktur untuk melakukan data mining. Kerangka kerja ini terdiri dari tujuh langkah, yang dapat membantu peneliti untuk melakukan data mining secara sistematis dan terstruktur.

Indeks Davies-Bouldin (DBI) adalah metode evaluasi hasil data mining yang digunakan untuk mengukur homogenitas dan heterogenitas antar kelompok. Software pembantu, seperti rapidminer versi 9.10, dapat memudahkan proses data mining dengan menyediakan berbagai fitur dan tools yang dibutuhkan.

**1. Kerangka Penelitian**



Sumber: Olahan Peneliti (2023)  
Gambar 2. Kerangka Penelitian

1. Dataset, data yang akan diolah untuk mencapai tujuan penelitian.
2. Penentuan Variabel, menentukan variabel yang akan digunakan dan yang tidak akan digunakan.
3. Persiapan data, melakukan *cleansing* serta *transform* agar data siap untuk masuk proses tahapan data mining.
4. Filtering data, melakukan analisa data untuk mencari *insight* dan *trends* pada dataset.
5. Proses ke dalam rapidminer, memasukkan dataset yang sudah dipersiapkan ke *repository* rapidminer untuk diolah.
6. Penerapan algoritma, dataset akan diterapkan algoritma k-means dan menampilkan data yang siap untuk di analisa.
7. Hasil dan analisa, data yang sudah memiliki *output* akan di analisa serta dipresentasikan.

**2. CRISP-DM**

CRISP-DM merupakan metodologi yang mempersiapkan sebuah proses standar de-facto terstruktur dalam perencanaan proyek data mining. Proses ini juga dapat diterapkan dari berbagai sektor *industry*, sehingga CRISP-DM merupakan model proses ter-organisasi dan tidak terbatas pada teknologi apa pun. Selain itu model proses independen industri

untuk penambahan data terdiri dari enam fase iteratif dari *business understanding* hingga *deployment* yang secara garis besar menggambarkan gagasan utama, tugas dan hasil keluaran (Schröer et al., 2021)



Sumber: CRISP-DM Metode (2020)  
Gambar 3. Fase dan tahapan CRIPS-DM

**3. Klaster**

Klasterisasi adalah teknik penambangan data yang membagi data ke dalam kelompok berdasarkan kemiripan. Proses ini dapat digunakan untuk memahami perbedaan dan persamaan antara kelompok data (Ignatius Moses Setiadi et al., 2020).

Algoritma pengelompokan konvensional tidak diskalakan dengan kumpulan data besar. Untuk mengatasinya, pendekatan praktis adalah paralelisasi algoritma (Praveen & Jayanth Babu, 2019)

**4. K-means**

K-Means adalah metode partisi yang iteratif untuk membagi data ke dalam kelompok-kelompok yang berbeda dengan meminimalkan rata-rata jarak setiap data ke klasternya. K-Means akan menghasilkan titik *centroid*, lalu setiap objek dikelompokkan ke dalam klaster berdasarkan jarak terdekat ke *centroid*. (Sembiring et al., 2020)

Nilai titik pertama dipilih secara *random*, lalu dihitung jaraknya ke semua titik data. Data yang jaraknya dekat membentuk klaster. Proses ini akan terus berlanjut sampai tidak ada perubahan (Alam et al., 2019), Untuk cara perhitungan Euclidean Distance:

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

Keterangan:

- D = Lambang *Euclidean Distance*
- X = Banyaknya objek
- ∑ p= Jumlah Data yang akan di record

Algoritma K-means memiliki waktu eksekusi yang cepat, mudah diimplementasikan, dan dapat mengurangi kompleksitas data (Umargono et al., 2020).

Namun menurut Ahmed, M pada tahun 2020, algoritma k-means memiliki beberapa kelemahan seperti:

1. Centroid awal ditentukan secara acak
2. Jumlah kluster (K) harus ditentukan dengan tepat
3. K-Means tidak dapat menginisialisasi *centroid* secara optimal.

Maka digunakanlah Davies-Bouldin Indeks(DBI) untuk mengukur validitas kluster. DBI bertujuan untuk memaksimalkan jarak antar kluster dan meminimalkan jarak intra-kluster. Dengan cara ini, perbedaan antar kluster menjadi jelas dan setiap objek dalam kluster memiliki kesamaan karakteristik yang tinggi. DBI dapat digunakan untuk mengevaluasi kualitas kluster yang dihasilkan oleh k-means. perhitungan DBI menurut Davies dan Bouldin adalah sebagai berikut:

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i) \dots\dots\dots(2)$$

Keterangan:

SSW = Sum of Squre within cluster

$m_i$  = Akumulasi hasil data dalam kluster ke-i

$c_i$  = Centroid kluster ke-i

**III. HASIL DAN PEMBAHASAN**

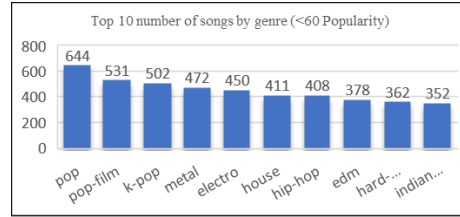
**1. Metode Pemilihan Populasi dan Sampel**

*Business understanding* pada penelitian kali ini untuk kluster spotify sebagai berikut:

- a. Dataset memiliki 113.999 data lagu yang siap untuk dikluster fitur atribut audionya.
- b. Kumpulan data lagu Spotify tersebut memiliki atribut yang diskalakan menggunakan angka.
- c. Adanya faktor-faktor tertentu yang dimiliki atribut dalam melakukan penerapan data mining.

Sampel data sebanyak 14.822 lagu dengan *popularity* di atas 60 akan diuji untuk atribut *danceability, energy, loudness, speechiness, acousticness, instrumentalness, liveness, valance, tempo dan track\_genre*. Track genre akan digunakan sebagai id.

Dengan mengetahui atribut apa saja yang akan digunakan maka peneliti mencoba untuk mengetahui genre yang paling banyak pada *popularity* 60 ke atas, serta memvisualisasikan, maka hasil yang dipadatkan bisa dilihat diagram berikut :

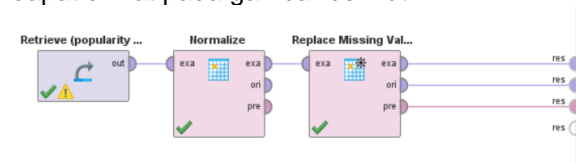


Sumber: Olahan Peneliti (2023)

Gambar 4. 10 Genre dengan lagu terbanyak

Pada diagram diatas bisa dilihat terdapat 10 genre yang memiliki jumlah lagu terbanyak dari data testing. Genre yang dimiliki seperti *pop, pop-film, k-pop, metal, electro, house, hip-hop, edm, hard-rock* dan *indian-pop*. Genre *pop* menempati posisi teratas dengan memiliki jumlah lagu sebanyak 644, dan yang paling terakhir adalah genre *indian-pop* dengan memiliki jumlah lagu sebanyak 352 pada diagram *Top 10 number of songs by genre*.

Persiapan data pada rapidminer menggunakan operator *retrieve* untuk membaca dataset dan operator *normalize* untuk menormalkan data. Normalisasi menggunakan transformasi-z untuk menghasilkan rata-rata 0 dan standar deviasi 1 untuk setiap atribut. Dengan normalisasi, semua atribut berada pada skala yang sama dan dapat dibandingkan satu sama lain. Untuk contoh pengoperasian dengan rapidminer dapat dilihat pada gambar berikut:

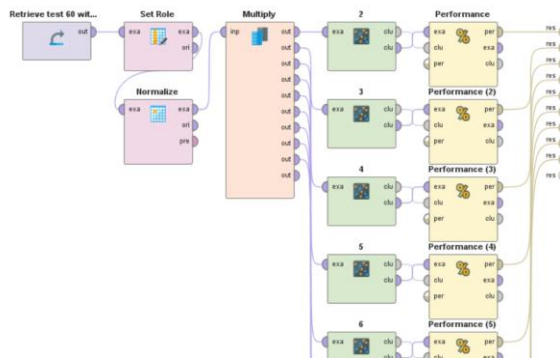


Sumber: Olahan Peneliti (2023)

Gambar 5. Rapidminer *preparation data*

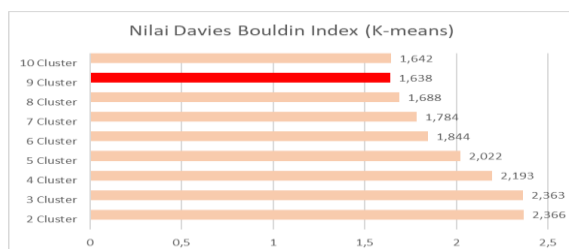
**2. Eksperimen dan Pengujian Algoritma**

Tahapan *modeling* merupakan tahapan berikutnya dari *data preparation*. peneliti akan menerapkan algoritma k-means, *multiply operator*, dan *operator performance* untuk menghitung hasil kluster K-means dan menganalisa nilai *davies Bouldin Indeks*. Pengujian pada fase modeling akan melakukan perhitungan klustering dari jumlah 2 kluster hingga 10 kluster. Jumlah kluster yang paling optimal akan dilihat dari nilai *davies bouldin indeks*. Untuk contoh pemodelan k-means dalam rapidminer bisa dilihat pada gambar berikut:



Sumber: Olahan Peneliti (2023)  
Gambar 6. Pemodelan K-means

Setelah melakukan proses pemodelan di atas, akan dikumpulkan dari hasil perhitungan dari setiap kluster dan dibandingkan hasilnya untuk mencari nilai terkecil dari setiap hasil proses yang ada untuk dilanjutkan untuk pengujian dengan auto model sekaligus melakukan tahap *deployment*. hasil dari DBI pada masing-masing pengujian kluster dapat dilihat tabel berikut:



Sumber: Olahan Peneliti (2023)  
Gambar 7. Hasil DBI

Berdasarkan hasil DBI dari percobaan yang dilakukan, dengan mencoba menerapkan jumlah kluster yang berbeda, ternyata kluster terbaik berada pada percobaan ke-8 yaitu perhitungan dengan jumlah kluster 9 memiliki nilai DBI 1,638. Menandakan bahwa dalam penelitian mendapatkan nilai jumlah kluster yang optimal berada di kluster 9, dan memiliki *Avg.within centroid distance* sebesar 5.006 dengan memperoleh nilai *centroid distance* pada setiap klusternya sebagai berikut:

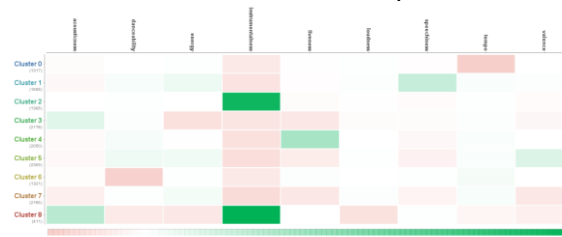
Tabel 2. Nilai rata-rata dari setiap kluster

<i>Avg. within centroid distance_cluster</i>	
Jumlah Kluster	Value
2	7.374
3	6.115
4	4.788
5	6.413
6	3.501
7	4.974
8	9.089
9	3.331
10	4.693

Sumber: Olahan Peneliti (2023)

### 3. Hasil Akhir

Pada proses *deployment*, peneliti akan menampilkan visualisasi hasil perhitungan dan hasil penelitian. *Heat map* adalah metode visualisasi data yang digunakan untuk menggambarkan hubungan antara dua atau lebih variabel. *Heat map* menggunakan warna untuk mewakili nilai data, dengan warna yang lebih terang mewakili nilai yang lebih tinggi. Berikut adalah hasil dari *heat map*:



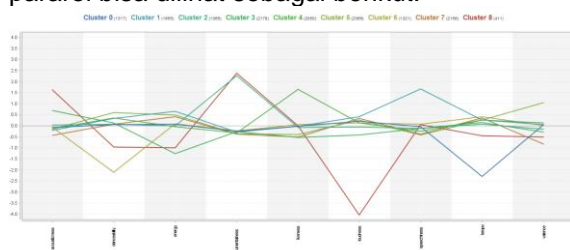
Sumber: Olahan Peneliti (2023)  
Gambar 8. Heat Map

Menjelaskan sekaligus mengidentifikasi atribut yang paling penting pada setiap kluster. Untuk penjelasan lebih jelasnya akan didefinisikan seperti berikut:

- Pada kluster 0 memiliki jumlah anggota 1317, yang memiliki atribut tempo di rata-rata 90,15% lebih kecil, *instrumentalness* berada pada rata-rata 42.49% lebih kecil dan *acousticness* berada pada rata-rata 6.09% lebih kecil.
- Pada kluster 1 memiliki jumlah anggota 1658, yang memiliki atribut *speechiness* di rata-rata 94,38% lebih besar, *instrumentalness* berada pada rata-rata 51.99% lebih kecil dan *energy* berada pada rata-rata 28.79% lebih besar.
- Pada kluster 2 memiliki jumlah anggota 1365, yang memiliki atribut *instrumentalness* di rata-rata 392.06% lebih besar, *Valance* berada pada rata-rata 8.82% lebih kecil dan *speechiness* berada pada rata-rata 7.33% lebih besar.
- Pada kluster 3 memiliki jumlah anggota 2178, yang memiliki atribut *energy* di rata-rata 57.20% lebih kecil, *acousticness* berada pada rata-rata 48.10% lebih besar dan *instrumentalness* berada pada rata-rata 47.74% lebih kecil.
- Pada kluster 4 memiliki jumlah anggota 2050, yang memiliki atribut *liveness* di rata-rata 142.05% lebih besar, *instrumentalness* berada pada rata-rata 56.04% lebih kecil dan *speechiness* berada pada rata-rata 14.07% lebih kecil.
- Pada kluster 5 memiliki jumlah anggota 2369, yang memiliki atribut *instrumentalness* di rata-rata 61.88% lebih

- kecil, *valance* berada pada rata-rata 57.58% lebih besar dan *liveness* berada pada rata-rata 34.76% lebih kecil.
- g. Pada kluster 6 memiliki jumlah anggota 1321, yang memiliki atribut *energy* di rata-rata 85.27% lebih kecil, *instrumentalness* berada pada rata-rata 41.03% lebih kecil dan tempo berada pada rata-rata 15.99% lebih besar.
  - h. Pada kluster 7 memiliki jumlah anggota 2156, yang memiliki atribut *instrumentalness* di rata-rata 66.36% lebih kecil, *valance* berada pada rata-rata 45.84% lebih kecil dan *liveness* berada pada rata-rata 45.12% lebih kecil.
  - i. Pada kluster 8 memiliki jumlah anggota 411, yang memiliki atribut *instrumentalness* di rata-rata 412.55% lebih besar, *acousticness* berada pada rata-rata 113.86% lebih besar dan *loudness* berada pada rata-rata 55.13% lebih kecil.

Dengan menganalisa hasil di atas atribut *instrumentalness* merupakan atribut yang paling terlihat kontras dibanding dengan atribut lainnya. Apabila dilihat pada kluster 3 dan kluster 9, atribut *intrumentalness* adalah yang terpenting sedangkan pada kluster lainnya *instrumentalness* menunjukkan *value* yang kurang bagus ditandai dengan warna merah. Pada atribut *loudness* terlihat semua berwarna putih pada setiap klasternya kecuali pada kluster 9 yang menunjukkan warna merah, yang menandakan bahwa *loudness* memiliki peran dan keterkaitan yang serupa pada setiap kluster yang ada. *Centroid chart*, menunjukkan nilai-nilai untuk *cluster centroids* dalam bentuk bagan paralel dan tabel agar lebih mudah dipahami untuk di analisa, hasil *centroid chart* bentuk bagan paralel bisa dilihat sebagai berikut:



Sumber: Olahan Peneliti (2023)  
Gambar 9. Centroid chart K-means

*Scatter plot* digunakan dalam pengelompokan untuk mengidentifikasi pola atau kelompok dalam data. Salah satu contoh *scatter plot* pada penelitian ini terdapat pada gambar berikut.:



Sumber: Olahan Peneliti (2023)  
Gambar 9. Scatter plot kluster 1

*Scatter plot* pada kluster satu memiliki variabel *intrumentalness* dan tempo, untuk titik-titik lebih banyak berkumpul ke arah tempo. Apabila melihat dari seluruh hasil *scatter plot* dari kluster 1 sampai 9 maka bisa dibentuk ke dalam tabel seperti berikut:

**Tabel 3. Hasi seluruh *scatter plot*(SP)**

Atribut	Kemunculan di SP	Dominan
<i>Acousticness</i>	2	1
<i>Danceability</i>	1	1
<i>Energy</i>	1	1
<i>Instrumentalness</i>	8	0
<i>Liveness</i>	1	1
<i>Loudness</i>	0	0
<i>Speechiness</i>	1	1
<i>Tempo</i>	1	1
<i>Valance</i>	3	3
<b>Total</b>	<b>18</b>	<b>9</b>

Sumber: Olahan Peneliti (2023)

**IV. KESIMPULAN**

Berdasarkan temuan dan uraian dalam penelitian yang dilakukan, peneliti akan menyajikan beberapa kesimpulan dan saran terkait dengan masalah penelitian. Berikut ini adalah ringkasan dari kesimpulan yang dapat diambil:

Atribut fitur audio yang selalu ada pada lagu-lagu populer terhadap lagu yang memiliki popularity di atas 60, adalah atribut *Instrumentalnes* serta atribut *valance*. Dengan kata lain, lagu-lagu populer di aplikasi spotify biasanya memiliki *insturmentalness* dan memiliki *valance*. Atribut *loudness* adalah atribut yang paling seimbang.

Penentuan jumlah kluster yang paling optimal untuk diterapkan dalam perhitungan, dalam pengelompokan ini, terdapat 9 kluster. Kluster ke-1 memiliki 1317 anggota, kluster ke-2 memiliki 1655 anggota, kluster ke-3 memiliki 1365 anggota, kluster ke-4 memiliki 2178 anggota, kluster ke-5 memiliki 2050 anggota, kluster ke-6 memiliki 2369 anggota, kluster ke-7 memiliki 1321 anggota, kluster ke-8 memiliki 2156 anggota, dan kluster ke-9 memiliki 411 anggota.

Penentuan pengelompokan dengan 9 klaster sudah dilakukan evaluasi dengan menganalisa hasil *davies bouldin indeks* memiliki nilai 1,638. Dalam dataset yang digunakan, genre pop memiliki lagu yang paling banyak untuk *popularity* diatas 60 dengan jumlah lagu sebanyak 644.

Saran untuk penelitian ini memiliki banyak potensi, seperti menggunakan data Spotify terbaru, memisahkan sub-genre dan genre utama, mencoba operator rapidminer yang lebih banyak, membahas variabel lain selain fitur atribut audio, mengklasifikasi lirik lagu yang *explicit*, dan mencari ruang lingkup yang lebih luas.

## V. REFERENSI

- Alam, M. S., Rahman, M. M., Hossain, M. A., Islam, M. K., Ahmed, K. M., Ahmed, K. T., Singh, B. C., & Miah, M. S. (2019). Automatic human brain tumor detection in mri image using template-based k means and improved fuzzy c means clustering algorithm. *Big Data and Cognitive Computi* 3(2), 1–18. <https://doi.org/10.3390/bdcc3020027>
- Dewatara, G. W., & Agustin, S. M. (2019). Pemasaran Musik Pada Era Digital Digitalisasi Industri Musik Dalam Industri 4.0 Di Indonesia. *WACANA, Jurnal Ilmiah Ilmu Komunikasi*, 18(1). <https://doi.org/10.32509/wacana.v18i1.729>
- Florez Ramos, E., & Blind, K. (2020). Data portability effects on data-driven innovation of online platforms: Analyzing Spotify. *Telecommunications Policy*, 44(9), 102026. <https://doi.org/10.1016/j.telpol.2020.102026>
- Ignatius Moses Setiadi, D. R., Satriya Rahardwika, D., Rachmawanto, E. H., Atika Sari, C., Irawan, C., Kusumaningrum, D. P., Nuri, & Trusthi, S. L. (2020). Comparison of SVM, KNN, and NB Classifier for Genre Music Classification based on Metadata. *Proceedings - 2020 International Seminar on Application for Technology of Information and Communication: IT Challenges for Sustainability, Scalability, and Security in the Age of Digital Disruption, ISemantic 2020*, 12–16. <https://doi.org/10.1109/iSemantic50169.2020.9234199>
- I Made Laut Mertha Jaya. (2020). *Metode Penelitian Kuantitatif dan Kualitatif: Teori, Penerapan, dan Riset Nyata* (p. 232). Fira Husaini. [https://books.google.co.id/books/about/Method\\_Penelitian\\_Kuantitatif\\_dan\\_Kualitatif?id=yz8KEAAAQBAJ&redir\\_esc=y](https://books.google.co.id/books/about/Method_Penelitian_Kuantitatif_dan_Kualitatif?id=yz8KEAAAQBAJ&redir_esc=y)
- Jollyta, D. W. R. M. Z. (n.d.). *Konsep data mining dan penerapan.pdf*.
- Mukhid, A. (2021). Metodologi Penelitian Pendekatan Kuantitatif. In *Syria Studies* (Vol. 7, Issue 1). [https://www.researchgate.net/publication/269107473\\_What\\_is\\_governance/link/548173090cf22525dcb61443/download%0Ahttp://www.econ.upf.edu/~reynal/Civilwars\\_12December2010.pdf%0Ahttps://thinks-asia.org/handle/11540/8282%0Ahttps://www.jstor.org/stable/41857625](https://www.researchgate.net/publication/269107473_What_is_governance/link/548173090cf22525dcb61443/download%0Ahttp://www.econ.upf.edu/~reynal/Civilwars_12December2010.pdf%0Ahttps://thinks-asia.org/handle/11540/8282%0Ahttps://www.jstor.org/stable/41857625)
- Pandya, M. (2023). Spotify tracks dataset. Kaggle. Diakses pada 23 Maret 2023. <https://www.kaggle.com/datasets/maharshipandya/-spotify-tracks-dataset>
- Praveen, P., & Jayanth Babu, C. (2019). Big Data Clustering: Applying Conventional Data Mining Techniques in Big Data Environment. In *Lecture Notes in Networks and Systems* (Vol. 74). Springer Singapore. [https://doi.org/10.1007/978-981-13-7082-3\\_58](https://doi.org/10.1007/978-981-13-7082-3_58)
- Prey, R., Esteve Del Valle, M., & Zwerwer, L. (2022). Platform pop: disentangling Spotify's intermediary role in the music industry. *Information Communication and Society*, 25(1), 74–92. <https://doi.org/10.1080/1369118X.2020.1761859>
- Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181(2019), 526–534. <https://doi.org/10.1016/j.procs.2021.01.199>
- Sembiring, F., Octaviana, O., & Saepudin, S. (2020). Implementasi Metode K-Means Dalam Pengklasteran Daerah Pungutan Liar Di Kabupaten Sukabumi (Studi Kasus: Dinas Kependudukan Dan Pencatatan Sipil). *Jurnal Tekno Insentif*, 14(1), 40–47. <https://doi.org/10.36787/jti.v14i1.165>
- Umargono, E., Suseno, J. E., & S. K., V. G. (2020). *K-Means Clustering Optimization using the Elbow Method and Early Centroid Determination Based-on Mean and Median*. 474(Isstec 2019), 234–240. <https://doi.org/10.5220/0009908402340240>
- Uswatun, L. (2021). Metode Pengumpulan Data dalam Penelitian Kualitatif. *UNPAR Press*, 1(1), 1–29. <https://www.dqlab.id/metode->



pengumpulan-data-dalam-  
penelitian%0Ahttp://ciputrauceo.net/blog/2  
016/2/18/metode-pengumpulan-data-  
dalam-penelitian

Vonderau, P. (2019). The Spotify Effect: Digital  
Distribution and Financial Growth.  
*Television and New Media*, 20(1), 3–19.  
<https://doi.org/10.1177/1527476417741200>