

K-Means Analysis for Music Segmentation in Playlist Recommendation

Abdul Hamid¹, Syarief Abdullah², Galbin Adi Pradana³, and Diki Gita Purnama⁴

^{1,2,3,4}Paramadina University, Informatics Engineering, East Jakarta, Indonesia

*Corresponding author: abdul.hamid@students.paramadina.ac.id

ABSTRACT

Nowadays, music streaming platforms are the main choice for music lovers to enjoy songs digitally. To improve user experience, a recommendation system is needed that is able to suggest personalized music based on individual preferences. This research examines the application of the K-Means Clustering algorithm in audio feature-based music segmentation to generate musically and emotionally relevant playlists. Spotify was chosen as the sample in the study because it is the largest and most widely used music streaming platform in the world. The dataset is obtained from Kaggle and contains audio attributes such as tempo, energy, valence, and danceability. After preprocessing and normalizing the data, the K-Means algorithm was used to group songs into clusters based on the similarity of their audio features. The clustering produced four distinct song groups, each characterized by specific audio traits such as tempo, energy, and valence. These clusters were then mapped into playlist themes, such as Energetic & Uplifting and Mellow & Relaxing, based on their dominant features. In a limited user test involving ten listeners, the playlists were rated positively in terms of musical flow and emotional coherence. Notably, Cluster 2 received the highest satisfaction scores. These results indicate that clustering songs based on audio features is not only technically viable but also enhances the emotional alignment of music recommendations.

Keywords: K-Means Clustering, Music Recommendation, Song Segmentation, Streaming Music Playlist

1. INTRODUCTION

The rapid growth of digital music platforms has changed the way people access and experience music. Services such as Spotify have become essential in delivering personalized content using advanced recommendation algorithms (Jing et al., 2024). These systems are expected not only to match user preferences but also to adapt to emotional and contextual user states (Jing et al., 2024). Consequently, interest is increasing in building systems that combine technical performance with emotional intelligence (Melchiorre et al., 2023). While music recommendation systems have advanced, most still focus on individual song suggestions without considering playlist-level coherence. This often leads to playlists that feel disjointed or emotionally inconsistent. (Chan et al., 2024) emphasized the role of emotional and contextual alignment in improving user experience, while (Chen et al., 2024) showed that clustering based on audio features can enhance personalization. However, few studies explore how clustering can be used to form coherent playlists. This study addresses that gap by applying K-Means for music segmentation to improve emotional flow in playlist recommendations.

Earlier recommendation systems relied heavily on collaborative filtering, which is prone to cold-start and sparsity issues (Singh & Dembla, 2023). Newer systems now incorporate hybrid and content based models that use audio features like tempo, valence, danceability,

and acousticness to define song characteristics (Takashima et al., 2023). These features help machine learning algorithms like K-Means cluster songs based on similarity in audio and emotion (Chen et al., 2024). This approach allows playlists to be generated with a higher degree of musical and emotional coherence (Rohman & Wibowo, 2024).

K-Means remains a preferred clustering algorithm due to its simplicity and ability to manage high-dimensional data effectively (Farhani & Qoiriah, 2024). Several studies have successfully applied K-Means to segment Spotify tracks into clusters based on mood and acoustic similarity (Hakim et al., 2025). These models often involve preprocessing such as Min Max normalization and validation using tools like the Elbow Method, Silhouette Score, and Davies Bouldin Index (Ikhsan Firmansyah et al., 2023).

However, many of these systems are limited to technical evaluations, overlooking the user's emotional experience. Clusters that look good on paper may not match how listeners perceive and respond to music (Purba et al., 2020). To address this, recent research has integrated psychological models and affective computing into recommendation frameworks (Daffa Rachman & Voutama, 2024). These models aim to create emotionally aligned playlists that improve engagement and perceived relevance (Chan et al., 2024) (Marlia et al., 2024).

In Indonesia, several researchers have adopted K-Means for music clustering, especially for emotional segmentation. (Farhani & Qoiriah, 2024) demonstrated moderate clustering performance using Spotify audio features. (Ikhsan Firmansyah et al., 2023) showed how normalization affects clustering results. Spotify is often chosen as a case study due to its accessible metadata and global use (Hakim et al., 2025). (Daffa Rachman & Voutama, 2024) added user based validation to assess emotional fit.

This study builds on those efforts by applying K-Means to a dataset of 2,000 Spotify tracks. It explores how well unsupervised clustering can form playlists that feel emotionally consistent and satisfying to users. The clustering process uses nine audio-related attributes, such as tempo, energy, valence, and acousticness, as input parameters to group songs with similar musical characteristics. Although Spotify serves as the primary data source, the proposed method can be applied to other streaming platforms with similar audio feature availability.

The proposed method combines internal validation metrics like Silhouette Score and Davies Bouldin Index with qualitative assessments from listeners. This dual evaluation framework aims to produce playlists that are not only technically sound but also emotionally meaningful and contextually appropriate.

2. METHODS

This study employs a quantitative-experimental approach to cluster music tracks based on their intrinsic audio characteristics using the K-Means Clustering Algorithm. The overall workflow adopts the Cross Industry Standard Process for Data Mining (CRISP DM) methodology, which consists of six structured stages: understanding the objectives, data understanding, data preparation, modeling, evaluation, and deployment. This framework was chosen for its iterative and adaptive structure, widely recognized for its effectiveness in data centric research across industries, including music informatics.

All data processing, clustering, and visualization procedures in this study were carried out using Python 3.10, implemented within a Jupyter Notebook environment. Several well established libraries were utilized to support the analysis, including Scikit-Learn for machine learning tasks, Pandas and NumPy for data manipulation, and Matplotlib and Seaborn for data visualization. The dataset used for this research was obtained from a publicly available source on Kaggle, specifically from the Spotify data visualization project by Varun Saikanuri (Saikanuri, n.d.). This dataset includes audio related features such as tempo, energy, valence, danceability, loudness, acousticness, instrumentalness, liveness, and speechiness, which were used as input parameters for clustering.

2.1 Data Collection

This research utilizes a dataset consisting of 2,000 music tracks sourced from Spotify via the Kaggle platform. Each entry contains both metadata, such as song title, artist, release year, and genre, and a set of numerical audio attributes. These include tempo, energy, danceability, valence, loudness, acousticness, speechiness, instrumentalness, and liveness, which serve as the foundation for the clustering process. The structure of the dataset and how these features are formatted can be seen in Figure 1.

	artist	song	duration_ms	explicit	year	
0	Britney Spears	Oops!...I Did It Again	211160	False	2000	
1	blink-182	All The Small Things	167066	False	1999	
2	Faith Hill	Breathe	250546	False	1999	
3	Bon Jovi	It's My Life	224493	False	2000	
4	*NSYNC	Bye Bye Bye	200560	False	2000	

	popularity	danceability	energy	key	loudness	mode	speechiness	
0	77	0.751	0.834	1	-5.444	0	0.0437	
1	79	0.434	0.897	0	-4.918	1	0.0488	
2	66	0.529	0.496	7	-9.007	1	0.0290	
3	78	0.551	0.913	0	-4.063	0	0.0466	
4	65	0.614	0.928	8	-4.806	0	0.0516	

	acousticness	instrumentalness	liveness	valence	tempo	genre
0	0.3000	0.000018	0.3550	0.894	95.053	pop
1	0.0103	0.000000	0.6120	0.684	148.726	rock, pop
2	0.1730	0.000000	0.2510	0.278	136.859	pop, country
3	0.0263	0.000013	0.3470	0.544	119.992	rock, metal
4	0.0408	0.001040	0.0845	0.879	172.656	pop

Figure 1. Sample of dataset structure used in the clustering process.

2.2 Data Preprocessing

To ensure reliable modeling, the dataset underwent several preprocessing steps. These included checking for and handling missing or duplicate entries, normalizing feature values, and removing irrelevant attributes. A Min-Max normalization technique was used to rescale all numeric values between 0 and 1, ensuring a balanced contribution from each feature in the clustering process. The steps of this process are outlined in Figure 2, which shows the preprocessing workflow.

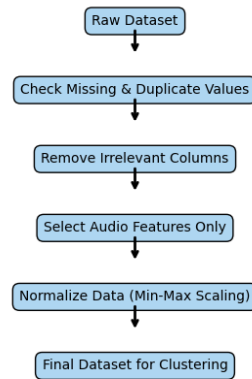


Figure 2. Workflow diagram of the preprocessing phase.

2.3 Feature Selection

From the available features, nine numerical audio-related attributes were selected to serve as inputs for clustering: tempo, energy, danceability, valence, acousticness, loudness, speechiness, instrumentality, and liveness. Categorical fields like genre, artist, and explicit content were excluded to maintain numerical consistency. These features were selected based on previous studies (Marlia et al., 2024) (Farhani & Qoiriah, 2024), which demonstrated their relevance in representing both musical characteristics and emotional nuance.

2.4 Clustering Using K-Means

The core modeling phase employed the K-Means algorithm, a widely used partitioning technique that assigns data points into k clusters based on feature similarity. The optimal number of clusters (k) was determined using two internal validation techniques: The Elbow Method, which examines the Within Cluster Sum of Squares (WCSS) across different k values (Figure 3), The Silhouette Coefficient, which measures how similar each data point is to its own cluster compared to others (Figure 4).

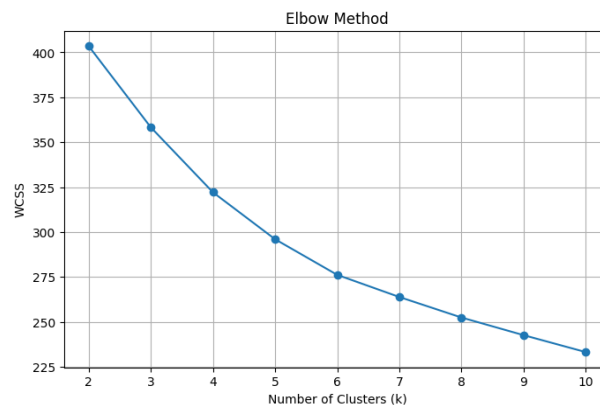


Figure 3. Elbow Method curve to determine optimal k .



Figure 4. Silhouette Score comparison for various k values.

The modeling was implemented in Python 3.10, within a Jupyter Notebook environment, using libraries such as Scikit Learn, Pandas, NumPy, and Matplotlib for computation and visualization.

2.5 Cluster Evaluation

To evaluate the performance of clustering, several techniques were employed. The Silhouette Score assesses how tightly grouped the points in each cluster are, while the Davies Bouldin Index provides a ratio based measure of intra cluster similarity and inter cluster separation. In addition to these numerical metrics, radar charts and scatter plots were used to visually interpret feature distributions across clusters. Figure 5 displays the radar plot of average feature values for each cluster, and Figure 6 provides a two dimensional scatter plot for visualizing cluster separation. A summary of evaluation scores is provided in Table 1.

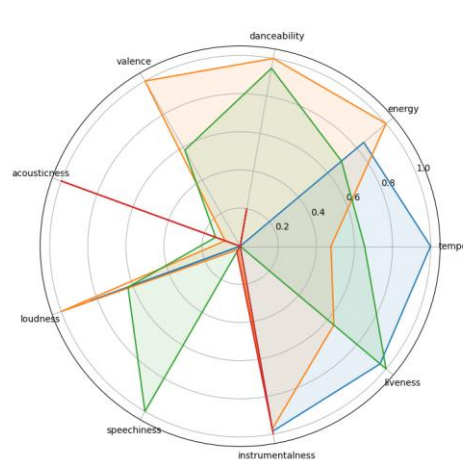


Figure 5. Radar plot showing feature averages per cluster.

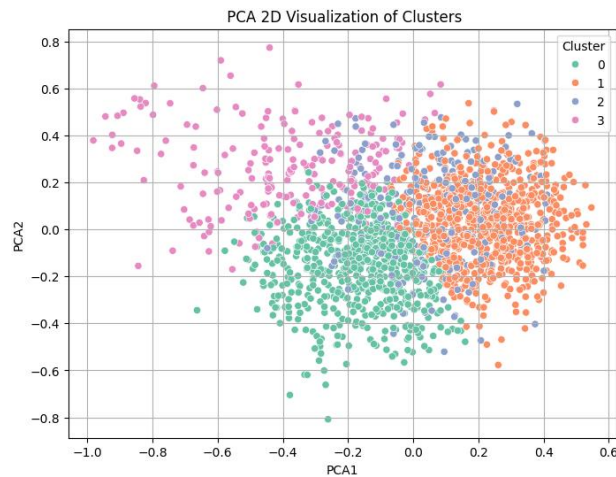


Figure 6. 2D scatter plot visualizing cluster distribution.

Table 1. Summary of clustering evaluation metrics.

Index	Metric	Value
0	Silhouette Score	0.1889
1	Davies Bouldin Index	1.671

2.6 Playlist Formation

Following the clustering process, each group was interpreted as a distinct playlist theme based on its average audio features. For example, a cluster characterized by high tempo and energy was associated with an "Energetic and Uplifting" playlist, while clusters with low valence and high acousticness were mapped to a "Mellow and Relaxing" theme. These labels aim to enhance personalization in music recommendations by aligning clusters with listening contexts.

2.7 User-Based Evaluation

To assess the subjective quality of playlists generated from the clusters, a user study involving ten participants was conducted. Each participant was asked to rate the coherence, emotional alignment, and satisfaction of the playlists on a five point Likert scale. The user feedback was later analyzed to provide qualitative validation of the clustering process.

Average User Rating per Cluster				
Cluster	Musical Flow	Emotional Coherence	Satisfaction	
0	4.2	4.0	4.1	
1	3.9	3.7	3.8	
2	4.5	4.4	4.6	
3	4.1	4.0	4.2	

Figure 7. User rating results on playlist coherence

Tools and Environment

Programming Language: Python 3.10

Libraries: Scikit-Learn, Pandas, NumPy, Matplotlib, Seaborn

Environment: Jupyter Notebook

Dataset Source:

<https://www.kaggle.com/code/varunsaikanuri/spotify-data-visualization>

3. RESULTS AND DISCUSSIONS

This section presents the findings from the clustering process, performance evaluation, playlist formation, and supporting user feedback. Each result is analyzed based on the methods previously described to assess the effectiveness of the K-Means approach in segmenting songs for personalized playlist creation.

3.1 Clustering Results

After executing the K-Means clustering algorithm with the optimal number of clusters ($k = 4$), the dataset was segmented into four distinct groups based on similarities in nine audio features.

Figure 8 presents a comprehensive summary that outlines how each cluster differs in its audio profile and suggests an appropriate playlist theme based on these characteristics.

Cluster	tempo	energy	danceability	valence	acousticness	loudness	speechiness	instrumentalness	liveness	Playlist Theme
0	125.882	0.736	0.589	0.371	0.06	-5.358	0.068	0.017	0.197	Mixed Mood
1	117.822	0.778	0.729	0.744	0.094	-4.932	0.073	0.017	0.175	Energetic and Uplifting
2	120.54	0.694	0.722	0.587	0.116	-5.898	0.294	0.006	0.2	Mixed Mood
3	110.471	0.503	0.617	0.369	0.472	-7.523	0.076	0.017	0.131	Emotional Ballads

Figure 8. Summary of average audio features across each cluster.

3.2 Evaluation of Clustering Performance

To evaluate the clustering results, two internal validation metrics were used: Silhouette Score and Davies Bouldin Index. As summarized in Table 1, the Silhouette Score obtained was 0.1889, indicating moderate cohesion within clusters. Meanwhile, the Davies Bouldin Index was 1.671, suggesting that while the clusters are distinguishable, their separation could be improved.

These metrics are visualized in Figure 9, which compares the evaluation results across different values of k . The results support the choice of $k = 4$, although the relatively low Silhouette Score highlights opportunities for enhancing cluster quality through algorithm optimization or feature enrichment.

	k	Silhouette Score	Davies-Bouldin Index
0	2	0.1882	1.9299
1	3	0.1820	1.8008
2	4	0.1889	1.6710
3	5	0.1862	1.5762
4	6	0.1704	1.6143
5	7	0.1604	1.6398
6	8	0.1316	1.7020
7	9	0.1327	1.6660
8	10	0.1364	1.6021

Figure 9. Clustering evaluation metrics for different values of k.

3.3 Playlist Formation Based on Clusters

Each cluster generated from the K-Means process was translated into a playlist based on its average audio feature profile. For instance, clusters with high tempo and energy were labeled as "Energetic & Uplifting", while those with low valence and high acousticness became "Mellow & Relaxing".

	Cluster	artist	song	Theme
0	0	Klangkarussell	Sonnentanz - Sun Don't Shine	Mixed Mood
1	0	Backstreet Boys	Shape of My Heart	Mixed Mood
2	0	Kanye West	Can't Tell Me Nothing	Mixed Mood
3	1	Eiffel 65	Move Your Body - Gabry Ponte Original Radio Edit	Energetic and Uplifting
4	1	Dominic Fike	3 Nights	Energetic and Uplifting
5	1	Gym Class Heroes	Ass Back Home (feat. Neon Hitch)	Energetic and Uplifting
6	2	Fedde Le Grand	Put Your Hands Up For Detroit - Radio Edit	Mixed Mood
7	2	NLE Choppa	Shotta Flow (feat. Blueface) [Remix]	Mixed Mood
8	2	A\$AP Ferg	Plain Jane	Mixed Mood
9	3	Jaheim	Put That Woman First	Emotional Ballads
10	3	Drake	Pop Style	Emotional Ballads
11	3	Beyoncé	Halo	Emotional Ballads

Figure 10. Playlist theme and sample songs from each cluster.

As illustrated in Figure 10, this mapping enables each playlist to reflect a distinct mood or context, supporting personalized music experiences based on acoustic similarity and emotional tone.

3.4 Discussion

As shown in Table 2, Cluster 2 received the highest ratings across all criteria, indicating that its playlist was the most emotionally engaging and musically cohesive. In contrast, Cluster 1 scored the lowest, suggesting weaker emotional fit and less satisfying flow. Clusters 0 and 3 received moderate feedback, showing decent but not outstanding results. These

differences may reflect how well the songs within each cluster align in terms of mood and acoustic features, with Cluster 2 likely offering a more consistent listening experience. This highlights the importance of not just grouping songs by similarity, but also considering their emotional coherence.

Table 2. Average user ratings on playlist evaluation criteria.

Index	Cluster	Musical Flow	Emotional Resonance	Satisfaction
0	0	4.3	4.4	4.2
1	1	4.0	3.9	3.8
2	2	4.5	4.6	4.7
3	3	4.1	4.0	4.1

Although the clustering achieved reasonable performance, the moderate evaluation scores suggest further improvements could be made. Future enhancements may involve testing alternative clustering methods, such as DBSCAN or hierarchical clustering, and incorporating lyrical or genre based features.

4. CONCLUSION

This study demonstrates the applicability of the K-Means Clustering algorithm for segmenting music tracks based on audio features, aiming to enhance the personalization of playlist recommendations. Using a dataset of 2,000 Spotify tracks and the CRISP DM framework, the research successfully grouped songs into clusters reflecting different musical and emotional profiles. The selected audio features such as tempo, energy, danceability, valence, and acousticness proved effective in capturing key aspects of song similarity relevant to listener preferences.

The clustering process yielded a Silhouette Score of 0.1889 and a Davies-Bouldin Index of 1.671, indicating moderate separation and cohesion across the clusters. The transformation of clusters into themed playlists, followed by positive user feedback on emotional alignment and musical flow, validated the practicality of the approach. Despite these promising results, the clustering quality suggests room for improvement. Future studies could explore alternative clustering techniques (for example DBSCAN or hierarchical methods), incorporate additional features such as lyrical content or genre tags, and expand user testing to include more diverse participant groups. These directions can further refine music segmentation and optimize personalized recommendation systems.

REFERENCES

Chan, J., Neveline, N., Sherina, S., Sinaga, F. M., & Halim, A. (2024). SISTEM REKOMENDASI MUSIK DENGAN MULTIMODAL FUSION ATTENTIVE

- NETWORKS. *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(1). <https://doi.org/10.23960/jitet.v12i1.3732>
- Chen, P., Dong, L., & Liu, Y. (2024). Design of music recommendation system based on EDA and K-means cluster analysis. *Applied and Computational Engineering*, 50(1), 317–323. <https://doi.org/10.54254/2755-2721/50/20241695>
- Daffa Rachman, M., & Voutama, A. (2024). IMPLEMENTASI ALGORITMA K-MEANS DALAM SISTEM REKOMENDASI MUSIK MENGGUNAKAN PYTHON. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 3857–3862. <https://doi.org/10.36040/jati.v8i3.9635>
- Farhani, F. S., & Qoiriah, A. (2024). *Klasterisasi Musik Menggunakan K-Means*. 06.
- Hakim, B., Kaunang, F. J., Susanto, C., Salim, J., & Indradjaja, R. (2025). IMPLEMENTASI MACHINE LEARNING DALAM PENGELOMPOKAN MUSIK MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING. *IDEALIS : InDonEsiA journal Information System*, 8(1), 74–83. <https://doi.org/10.36080/idealism.v8i1.3357>
- Ikhsan Firmansyah, M., Saepul Rohman, R., & 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>
- Jing, E., Liu, Y., Chai, Y., Yu, S., Liu, L., Jiang, Y., & Wang, Y. (2024). *Emotion-aware Personalized Music Recommendation with a Heterogeneity-aware Deep Bayesian Network* (Version 2). arXiv. <https://doi.org/10.48550/ARXIV.2406.14090>
- Marlia, S., Setiawan, K., & Juliane, C. (2024). Analysis of Music Features and Song Popularity Trends on Spotify Using K-Means and CRISP-DM. *SISTEMASI*, 13(2), 595. <https://doi.org/10.32520/stmsi.v13i2.3757>
- Melchiorre, A. B., Penz, D., Ganhör, C., Lesota, O., Fragoso, V., Fritzl, F., Parada-Cabaleiro, E., Schubert, F., & Schedl, M. (2023). Emotion-aware music tower blocks (EmoMTB): An intelligent audiovisual interface for music discovery and recommendation. *International Journal of Multimedia Information Retrieval*, 12(1). <https://doi.org/10.1007/s13735-023-00275-8>
- Purba, Y., Prayuda, J., Kom, S., Kom, M., & Kom, S. (2020). *PENERAPAN METODE K-MEANS CLUSTERING PADA DATA MINING UNTUK MENENTUKAN GENRE MUSIK LAGU DI RADIO JOY 101 FM*.
- Rohman, N., & Wibowo, A. (2024). CLUSTERING OF POPULAR SPOTIFY SONGS IN 2023 USING K-MEANS METHOD AND SILHOUETTE COEFFICIENT. *Jurnal Pilar Nusa Mandiri*, 20(1), 18–24. <https://doi.org/10.33480/pilar.v20i1.4937>
- Saikanuri, V. (n.d.). *Spotify Data Visualization*. <https://kaggle.com/code/varunsaikanuri/spotify-data-visualization>
- Singh, K. K., & Dembla, P. (2023). A Study on Emotion Analysis and Music Recommendation Using Transfer Learning. *Journal of Computer Science*, 19(6), 707–726. <https://doi.org/10.3844/jcssp.2023.707.726>
- Takashima, N., Li, F., Grzegorzec, M., & Shirahama, K. (2023). *Embedding-based Music Emotion Recognition Using Composite Loss* (No. arXiv:2112.07192). arXiv. <https://doi.org/10.48550/arXiv.2112.07192>