Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2609−2617
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9181

Hybrid Music Recommendation System Using K-Means Clustering and Neural Collaborative Filtering for Spotify Playlist Personalization

Rastomi Pamungkas[1,*], Permata[2], Rakhmat Dedi Gunawan[2], Adhie Thyo Priandika[2]

[1] Fakultas Teknik dan Ilmu Komputer, Informatika, Universitas Teknokrat Indonesia, Lampung, Indonesia
[2] Fakultas Teknik dan Ilmu Komputer, Magister Ilmu Komputer, Universitas Teknokrat Indonesia, Lampung, Indonesia
Email: [1]rastomi_pamungkas@teknokrat.ac.id, [2,*]permata@teknokrat.ac.id, [3]rakhmatdedig@teknokrat.ac.id, [4]adhie_thyo@teknokrat.ac.id
Correspondence Author Email: rastomi_pamungkas@teknokrat.ac.id
Submitted: **11/01/2026**; Accepted: **19/03/2026**; Published**: 19/03/2026**

Abstract−Personalizing music recommendations has become a significant challenge on music streaming platforms such as Spotify due to the vast number of available songs and the limitations of conventional recommendation systems in accurately capturing user preferences. In addition, traditional single-method recommendation approaches often face the cold start problem, which reduces the effectiveness of generated recommendations. Therefore, this study aims to develop and evaluate a hybrid recommendation system that integrates the K-Means Clustering algorithm and Deep Collaborative Filtering based on Neural Matrix Factorization to improve the relevance of music playlist recommendations. The dataset used in this study consists of more than 15,151 Spotify songs obtained from the Spotify dataset available on Kaggle. The dataset was processed through several stages including data inspection, data cleaning, feature selection, and standardization. Audio features used in the analysis include danceability, energy, acousticness, instrumentalness, valence, tempo, and duration. The optimal number of clusters was determined using the Elbow Method and Silhouette Score, resulting in five clusters with a relatively balanced data distribution. The clustering results were then used as the basis for Cluster-Based Filtering to narrow the search space of candidate songs before being processed by the Neural Matrix Factorization model. Performance evaluation was conducted using Hit Ratio at rank 10 and Normalized Discounted Cumulative Gain at rank 10. The proposed model achieved values of 0.1110 and 0.0507, respectively, indicating that the integration of clustering and deep collaborative filtering can improve the effectiveness and personalization of music recommendation systems. This study contributes by proposing a hybrid recommendation framework that integrates clustering-based item grouping with deep collaborative filtering to improve recommendation efficiency and playlist personalization in large-scale music streaming platforms.

Keywords: Music Recommendation; K-Means Clustering; Deep Collaborative Filtering; Neural Matrix Factorization; Spotify

1. INTRODUCTION

The growth of music streaming platforms such as Spotify, Apple Music, and JOOX in recent years has significantly transformed the way users access and enjoy music. As one of the largest platforms with millions of active users, Spotify faces significant challenges in providing accurate, relevant, and personalized music recommendations. The complexity of user preferences, the vast diversity of the music catalog, and the dynamics of listening habits make it difficult for traditional recommendation systems to deliver satisfactory results. The primary issue frequently encountered in music recommendation systems is the cold start problem, a condition where new users or new songs lack sufficient historical data for analysis. Additionally, data sparsity in interaction records leads to low prediction accuracy, while content-based approaches often fail to capture latent characteristics that represent the users' actual preferences. These conditions indicate the need for a more adaptive approach capable of integrating various information sources to enhance recommendation performance.

Efforts to address these issues have been explored in various recent studies. Deep learning-based recommendation systems have become widely adopted due to their ability to learn complex latent representations of users and items. Research on *Music Recommendation System on Spotify Using Deep Learning* [1] demonstrates that *deep neural network* methods can improve recommendation accuracy compared to traditional recommendation approaches. However, the study mainly focuses on deep learning models without incorporating techniques that can reduce the recommendation search space before the prediction process. Other research comparing *content-based*, *collaborative filtering*, and *hybrid filtering* approaches[2] found that the hybrid approach provides more stable and accurate results by leveraging the advantage of both approaches. Nevertheless, the study mainly combines conventional recommendation techniques and does not explore the integration of deep learning-based collaborative filtering models. Furthermore, research related to music clustering using the K-Means algorithm [3] shows that Spotify audio features such as *danceability*, *energy*, *valence*, and *tempo* can be used to form clear music segmentations, thereby helping to understand the characteristics of different music groups. However, the clustering results are primarily used for data grouping and are not integrated into a deep learning-based recommendation framework. These limitations indicated that previous studies tend to examine clustering techniques, hybrid filtering approaches, and deep learning models separately, leaving a research gap in integrating clustering-based item grouping with deep collaborative filtering models to improve the effectiveness and personalization of music recommendation systems.

Furthermore, research on boarding house recommendation systems based on Clustering and Collaborative Filtering [4] demonstrates that the integration of these two approaches can reduce the search space and improve recommendation accuracy. Although the application domain differs, the methodology of that study highlights the effectiveness of hybrid filtering strategies in recommendation systems. This approach is also relevant to the music domain because music recommendation systems must process a very large catalog of songs, making it important to

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2609−2617
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9181

limit the recommendation search space before performing prediction. By grouping similar items through clustering, the systems can focus on more relevant candidate songs, thereby improving recommendation efficiency and accuracy. Meanwhile, research on *Neural Collaborative Filtering (NCF)* for movie recommendations [5] proves that the *Neural Matrix Factorization (NeuMF)* model provides more accurate prediction results than conventional *matrix factorization* due to its ability to learn non-linear interactions between users and items. This capability is particularly important in music recommendation systems, where user listening behavior is often complex and influenced by multiple latent factors such as mood, genre preference, and listening context. Deep learning models such as NeuMF are therefore suitable for capturing these complex patterns in user-music interactions.

However, several limitations remain in existing studies. Research conducted in [6] investigates the use of the K-Means clustering algorithm to group Spotify songs based on audio features such as danceability, energy, and valence in order to identify music segments with similar characteristics. Although the study successfully forms several music clusters, the clustering results are mainly used for data grouping and are not integrated with deep learning-based recommendation models. Furthermore, research presented in [7] applies the Neural Matrix Factorization (NeuMF) model within the Neural Collaborative Filtering framework to improve recommendation accuracy by learning complex non-linear interactions between users and items. However, the implementation of NeuMF in that study is primarily conducted in domains such as movies, e-commerce, or hospitality, and its application in music recommendation systems that incorporate audio feature-based clustering remains limited. In addition, there is still limited research that explicitly uses K-Means clustering results as item groupings to restrict the recommendation search space before applying deep collaborative filtering models. In addition, previous studies using Spotify datasets such as [8] generally focus on descriptive analysis or conventional recommendation techniques using relatively moderate-sized datasets. In contrast, this research utilizes a Spotify dataset consisting of more than 15,000 songs with complete audio features, enabling a more comprehensive analysis of music characteristics and recommendation performance. Consequently, there is a research gap that has not been widely explored, namely the integration of unsupervised learning through clustering with deep learning-based recommendation models to significantly enhance playlist personalization.

Despite the advancements, several limitations remain in existing studies. Previous research generally investigates clustering techniques, hybrid filtering approaches, and deep learning models separately. Studies utilizing K-Means clustering mostly focus on music segmentation or descriptive analysis without integrating clustering results into the recommendation prediction process.On the other hand, studies applying Neural Collaborative Filtering models mainly rely on user-item interaction data without incorporating content-based grouping that can reduce the recommendation search space. Consequently, the integration of clustering-based item grouping with deep collaborative filtering models in music recommendation systems remains relatively limited. This research gap highlights the need for a hybrid framework that combines clustering-based item grouping with deep collaborative filtering to improve recommendation efficiency and personalization in large-scale music streaming platforms.

This research proposes a hybrid recommendation system that combines the K-Means Clustering algorithm with Deep Collaborative Filtering (DCF) based on Neural Matrix Factorization (NeuMF). In the initial stage, data is processed through cleaning, feature selection, and standardization of 15,151 songs from Spotify. Audio features such as danceability, energy, acousticness, instrumentalness, valence, tempo, and duration are used as the basis for the clustering process. Based on the Elbow method and Silhouette Score, five optimal clusters were obtained with distributions reflecting music types such as high-energy songs, acoustic songs, mellow songs, and instrumental tracks. These clustering results then serve as the initial search space to narrow down the recommendation area before being analyzed by the NeuMF model.

The objectives of this research are to: (1) develop a hybrid-based music recommendation system that integrates clustering and deep collaborative filtering; (2) analyze the quality of clusters formed from Spotify audio features; (3) evaluate the capability of the NeuMF model in predicting user preference scores; and (4) demonstrate that the integration of unsupervised learning methods with deep learning-based recommendations can mitigate the cold start problem and enhance music recommendation personalization on streaming platforms. Through this approach, this research is expected to provide both scientific and practical benefits in developing more adaptive, intelligent, and effective music recommendation systems.

The main contribution of this research lies in the integration of K-Means Clustering with Deep Collaborative Filtering based on Neural Matrix Factorization within a unified hybrid recommendation framework. Unlike previous studies that examine clustering or deep learning approaches separately, this study utilizes clustering results as a filtering mechanism to reduce the recommendation search space before applying the NeuMF model. This approach improves recommendation relevance while also increasing computational efficiency when dealing with large-scale music datasets.

2. RESEARCH METHODOLOGY

This research utilizes a Spotify song dataset encompassing over 15,000 tracks to develop a music recommendation system through a Hybrid Recommendation System approach. The steps include data collection as well as cleaning, selection, and standardization of audio features, song grouping using the K-Means Clustering algorithm, determination

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2609−2617
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9181

of the optimal number of clusters using the Elbow Method and Silhouette Score, plus the development of a recommendation model based on Deep Collaborative Filtering with Neural Matrix Factorization (NeuMF). Assessments are conducted to evaluate the clustering quality and the effectiveness of the generated suggestions. All stages are implemented using the Python programming language, with the support of Scikit-learn, TensorFlow, and Keras libraries. It is expected that these findings can improve music playlist personalization and address the cold start problem in music streaming services.

2.1 Research Stages

This research is conducted by utilizing a dataset obtained from the Spotify API as the primary data source. The study begins with the data collection and preparation stages, which serve as the foundation for the grouping process using the K-Means Clustering method. Before the data is processed for the clustering stage, *data Preprocessing* and *data validation* are performed to ensure the dataset does not contain dummy data, duplications, or invalid values. Subsequently, data cleaning, feature selection, and standardization are carried out on the audio features used, such as *danceability, energy, acousticness, instrumentalness, valence, tempo,* and *duration*. The processed data is then used for the training phase of the K-Means algorithm, with the optimal number of clusters determined using the Elbow Method and Silhouette Score. The results of the song clustering are then utilized as item groupings to narrow the recommendation search space before further processing using the Deep Collaborative Filtering method based on Neural Matrix Factorization (NeuMF). The final stage of the research involves an evaluation to analyze the clustering quality and the effectiveness of the resulting recommendation system. A more detailed illustration of these research stages can be found in Figure 1.

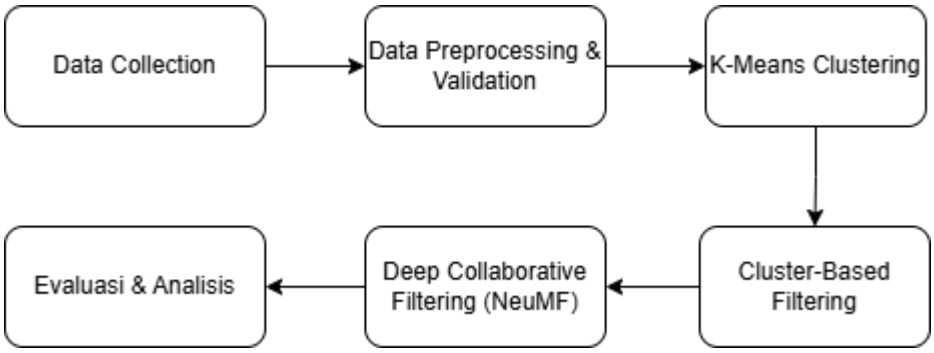

Figure 1. Research Stages

2.2 Data Collection

The research data was obtained from the Kaggle website in the form of a Microsoft Excel file containing more than 15,000 Spotify song entries. This dataset includes several key audio variables, such as danceability, energy, acousticness, instrumentalness, valence, tempo, and duration, which represent the characteristics of each song. This information also serves as the primary basis for the analysis, grouping, and development of the music recommendation system based on a Hybrid Recommendation System, which integrates K-Means Clustering and Deep Collaborative Filtering using Neural Matrix Factorization (NeuMF) [2].

2.3 Data Preprocessing & Validation

The Data Preprocessing and Validation stage is conducted to ensure the quality and suitability of the dataset before it is used in the analysis and modeling process. The dataset is first thoroughly examined to identify the data structure, data types, and value distribution of each audio feature [6]. At this stage, checks are performed for missing values, duplicate data, and value anomalies that fall outside logical ranges, particularly in numerical features such as *tempo, duration,* and *energy* [9].

Subsequently, the data validation process is carried out to ensure that all data used is authentic and representative, and does not contain dummy or synthetic data that could affect the research results. Validation is performed by checking the consistency of values across audio features, the alignment of value ranges with Spotify's standard audio characteristics, and the data distribution of each variable. Data detected as invalid, inconsistent, or duplicated is removed from the dataset. The output of this stage is a clean dataset that is suitable for the data preprocessing, clustering, and development phases of the Hybrid Recommendation System.

The results of the Data Preprocessing indicate that the dataset contains no significant duplicate data or missing values; therefore, all data is declared valid and feasible for the clustering and recommendation modeling processes.

2.4 K-Means Clustering

Clustering refers to the process of grouping entities, such as records, observations, or objects, based on their similarities to form classes or groups. One example of a clustering algorithm is K-Means Clustering. The K-Means Clustering algorithm operates by partitioning data into distinct groups based on the similarity of their attributes [8].

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2609−2617
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9181

The primary objective of this algorithm is to minimize intra-clustering variance while maximizing inter-cluster differences so that each group possesses homogeneous characteristics.

In this research, K-Means Clustering is applied to group Spotify songs based on audio features such as *danceability, energy, acousticness, instrumentalness, valence, tempo,* and *duration.* Before the clustering process is conducted, the data undergoes *data cleaning, feature selection,* and *standardization* to ensure scale uniformity across features [6]. The determination of the optimal number of clusters is performed using the *Elbow Method* and *Silhouette Score,* aiming to achieve a cluster count with the best balance between model complexity and grouping quality [10].

The evaluation results for the number of clusters are presented in the form of an *Elbow Method* graph and *Silhouette Score* values, as well as summarized in a distribution table of the number of songs in each cluster. These visualizations and tables are used to analyze the characteristics of each cluster and ensure that the resulting song segmentation has clear separation and representation. The generated clusters are then utilized as an initial stage (*pre-processing*) to narrow the Search Space before the data is further processed using the Deep Collaborative Filtering model.

2.4.1 Elbow Method and Silhouette Method

Determining the optimal number of clusters (K) is a crucial stage in the K-Means algorithm as it significantly influences the quality of the data grouping results. One commonly used approach is the Elbow Method, which involves calculating the *within-cluster sum of squares* (WCSS) for various K values. The WCSS value consistently decreases as the number of clusters increases; however, this decrease typically slows down at a specific point, forming an angle resembling an (Elbow). This point indicates the optimal number of clusters, as adding further clusters does not yield a significant improvement in cluster quality [11]. Nonetheless, determining the elbow point can often be subjective, especially with data containing complex patterns [12].

To address these limitations, the Silhouette Method is employed as a more objective supplementary evaluation method. This method measures cluster quality based on the degree of data cohesion and separation, using a *silhouette coefficient* ranging from -1 to 1. A value approaching 1 indicates that the data has high suitability within its cluster, which is considered favorable [12]. The optimal number of clusters is determined based on the highest average *silhouette coefficient* across the entire dataset [13].

The simultaneous use of the Elbow Method and Silhouette Method is considered capable of providing more valid results in determining the optimal number of clusters for the K-Means algorithm. Several previous studies have shown that the combination of these two methods is effective for grouping Spotify-based music data, both for recommendation systems and song characteristic analysis [13],[14]. Furthermore, cluster evaluation using the Silhouette Method yields values indicating good cluster quality, thereby supporting the selection of the K value obtained from the Elbow Method [15].

2.5 Cluster-Based Filtering

Cluster-Based Filtering is a filtering approach in recommendation systems that utilizes clustering results to narrow the *search space* before the recommendation process is conducted [4]. In this research, this approach is implemented using K-Means Clustering results, where each song is grouped based on the similarity of audio features such as *danceability, energy, acousticness, instrumentalness, valence, tempo,* and *duration* [8]. By limiting the recommendation process to relevant clusters, the system can improve computational efficiency and the relevance of the generated recommendations.

The output of *Cluster-Based Filtering* will subsequently be used as input for the *Deep Collaborative Filtering* based on *Neural Matrix Factorization* (*NeuMF*). This integration allows the model to learn user preference patterns more specifically within the scope of a particular cluster, thereby enhancing prediction accuracy and reducing *cold start* problems in music recommendation systems [2].

2.6 Deep Collaborative Filtering (NeuMF)

Deep Collaborative Filtering (DCF) is a recommendation system method that combines collaborative filtering with deep learning to capture non-linear relationships between users and items. In this research, DCF is implemented through *Neural Matrix Factorization* (NeuMF), which integrates two main elements: *Generalized Matrix Factorization* (GMF) and *Multi-Layer Perceptron* (MLP), with the aim of obtaining deeper latent representation of users and songs [16].

Mathematically, the GMF component models the linear interaction between users and songs through an element-wise product operation, formulated as:

$$\phi_{GMF} = p_u \odot q_i \tag{1}$$

Where p_u and q_i represent the latent vectors of the user and the song, respectively.

Meanwhile, the MLP component is used to capture non-linear relationships by combining the latent vectors of the user and the song, formulated as:

$$\phi_{MLP} = f(W[p_u; q_i] + b) \tag{2}$$

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2609−2617
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9181

With $\boldsymbol{f}(\cdot)$ as the non-linear activation function, W as the neural network weights and b as the bias.

The outputs from GMF and MLP are then combined to generate a predicted *preference score* using the following equation:

$$\widehat{y_{ui}} = \sigma(h^T[\phi_{GMF}; \phi_{MLP}]) \quad (3)$$

Where $\sigma(\cdot)$ is the sigmoid activation function and $\widehat{y_{ui}}$ indicates the preference level of user u toward song i.

The NeuMF model is trained using interaction data that has been filtered through the Cluster-Based Filtering stage, ensuring the learning process is focused on songs within relevant clusters. This approach is expected to improve recommendation accuracy and mitigate *cold start* problems in deep learning-based music recommendation systems [17].

2.7 Evaluation and Analysis

The evaluation and analysis of this model are conducted to assess the effectiveness of integrating *K-Means Clustering* and *Deep Collaborative Filtering* based *Neural Matrix Factorization (NeuMF)* in improving music recommendation quality. Evaluation results at the clustering stage indicate that the use of the Elbow Method and Silhouette Score successfully generates an optimal number of clusters that clearly represent song segmentation based on audio feature similarities, thereby supporting the Cluster-Based Filtering process within the recommendation system [18]. Furthermore, the performance analysis of the NeuMF model shows that learning user-song interaction patterns within a search space narrowed by clustering is able to enhance recommendation stability and relevance [19]. This hybrid approach is proven to be more effective than single-method approaches, particularly in mitigating the impact of *cold start* problems and improving the overall personalization level of music playlist recommendations, as also demonstrated in research on *hybrid* and *deep learning-based* recommendation systems within the domains of music and digital content [20].

3. RESULTS AND DISCUSSION

3.1 Data Collection

In the data collection stage, this research obtained a dataset from the Kaggle website, stored in an Excel file format and sourced from the Spotify Application Programming Interface (API). The dataset consists of 15,151 song entries containing several key audio variables, including danceability, energy, acousticness, instrumentalness, valence, tempo, and duration. The acquired data then serves as the foundation for the analysis and modeling of the music recommendation system, following examination, cleaning, and data preparation stages to ensure the quality and consistency of the dataset used in this study. An example of the Spotify song data structure used in this research is presented in Table 1.

Table 1. Spotify songs data

No	Track	Artist	Year	Danceability	Energy	Acousticness	Instrumentalness	Valence	Tempo
1	*Billie Jean*	Michael Jackson	1982	0.790	0.803	0.019	0.000	0.864	117.0
2	*Bohemian Rhapsody*	Queen	1975	0.392	0.402	0.271	0.000	0.224	72.7
3	*Imagine*	John Lennon	1971	0.605	0.313	0.826	0.000	0.379	75.0
4	*Hotel California*	Eagles	1976	0.585	0.508	0.296	0.000	0.609	147.0
5	*Smells Like Teen Spirit*	Nirvana	1991	0.521	0.954	0.000	0.000	0.720	116.0

3.2 Data Preprocessing & Validation

In the preprocessing stage, a *data cleaning* process is conducted to ensure the quality of the Spotify song dataset before further analysis. First, all audio attributes are examined to identify the presence of missing values, duplicate data, and anomalies. Subsequently, numerical features such as *danceability, energy, acousticness, instrumentalness, valence, tempo,* and *duration* are ensured to be in a consistent numerical format for computational processing. Invalid data, including missing values or duplicates, are removed from the dataset. Following this, a data standardization process is performed to equalize the scales across features, preventing any specific attribute from dominating the distance

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2609−2617
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9181

calculations in the K-Means algorithm. This stage produces a cleaner, more structured dataset that is ready for the clustering process and the development of the Hybrid Recommendation System.

3.3 K-Means Clustering

At this stage, based on the results of the clustering process on the dataset used, five main clusters were obtained with a relatively balanced data distribution. Cluster 1 has the largest membership with 5,384 songs, followed by Cluster 0 with 4,390 songs, Cluster 3 with 3,240 songs, Cluster 4 with 1,169 songs, and Cluster 2 with the smallest membership of 967 songs. This distribution indicates that the dataset possesses diverse variations in audio characteristics and is not concentrated in one specific group.
Subsequently, an analysis of the mean values for each audio feature within each cluster was conducted to identify the resulting music characteristics. A summary of the mean audio feature values for each cluster, along with the number of songs included, is presented in Table 2.

Table 2. Average clustering results

Cluster summary (mean values):

cluster	Year	Duration	Time_Signature	Danceability	Energy	Key	Loudness	Mode	Speechiness	Acousticness	Instrumentalness	Liveness	Valence	Tempo	Popularity
0	1997.617	252025.290	3.979	0.470	0.805	5.228	-6.253	0.666	0.070	0.084	0.101	0.250	0.420	129.680	47.281
1	1981.980	236655.340	3.998	0.685	0.669	5.330	-8.748	0.680	0.056	0.212	0.049	0.159	0.772	117.807	45.109
2	1974.791	241349.491	2.824	0.449	0.376	5.082	-12.092	0.784	0.047	0.604	0.118	0.196	0.419	125.807	35.722
3	1969.308	238594.657	4.018	0.523	0.333	5.092	-13.519	0.790	0.047	0.703	0.167	0.193	0.485	112.230	33.095
4	1999.705	229908.150	3.997	0.706	0.669	5.639	-7.394	0.559	0.293	0.182	0.016	0.223	0.577	119.088	51.018

Table 2 presents the average values of audio features for each cluster obtained from the K-Means clustering process. The results show distinct differences in musical characteristics across clusters. Cluster 4 has the highest danceability value (0.706) and relatively high energy (0.669), indicating that this cluster is dominated by rhythmic and energetic songs that are suitable for upbeat playlists. Cluster 1 also shows high danceability (0.685) and energy (0.669) with a relatively high valence value (0.772), suggesting that songs in this cluster tend to have a positive and lively mood. Cluster 3 is characterized by high acousticness (0.703) and relatively low energy (0.333), indicating that this cluster mostly contains acoustic or softer songs. Cluster 2 shows the highest acousticness value (0.604) with moderate energy levels, which suggests that this group may represent balanced songs between acoustic and rhythmic characteristics. Meanwhile, Cluster 0 shows moderate values across most features, indicating a more balanced mix of musical characteristics compared to other clusters.

3.3.1 Elbow Method and Silhouette Method

At this stage, the determination of the optimal number of clusters was conducted based on the evaluation results using the Elbow Method and Silhouette Method on the Spotify song dataset used in the study. Testing was performed by trying several *K* values to observe the resulting changes in *inertia* and *Silhouette Score*.

Based on the Elbow Method graph, a significant decrease in inertia is observed up to $K = 5$, after which the decline tends to level off at subsequent *K* values. This pattern indicates that $K = 5$ is the optimal point, as increasing the number of clusters beyond this value does not yield a meaningful improvement in clustering quality. The visualization of the Elbow Method and Silhouette Method results is shown in Figure 2.

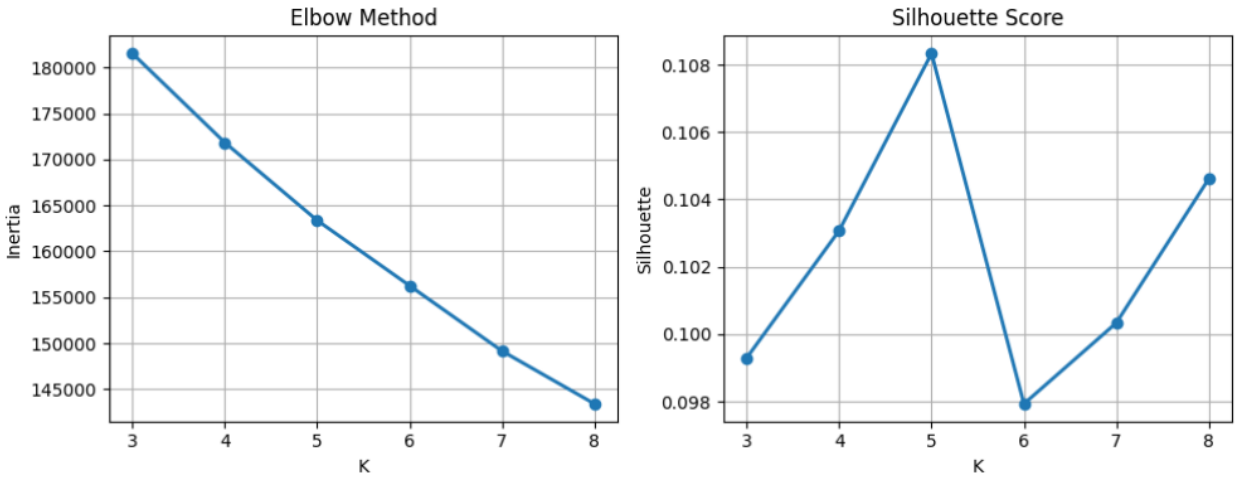

Figure 2. Elbow Method and Silhouette Method graphs

Figure 1 illustrates the evaluation process for determining the optimal number of clusters using the Elbow Method and Silhouette Score. The Elbow Method graph shows the relationship between the number of clusters (K)

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2609−2617
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9181

and the inertia value. As the number of clusters increases, the inertia value gradually decreases, indicating that the data points are grouped more compactly within each cluster. However, the decrease becomes less significant after K = 5, forming an elbow point that suggests an appropriate number of clusters.

The evaluation results using the Silhouette Method reinforce these findings. The highest Silhouette Score was obtained at K = 5, indicating that partitioning the data into five clusters produces the most effective and consistent level of separation between clusters compared to other K values. Based on the results of both evaluation methods, the number of clusters used in this study is set to five. These clustering results are subsequently utilized as the basis for implementing Cluster-Based Filtering to narrow the recommendation search space and improve recommendation relevance before further processing using the Deep Collaborative Filtering method.

3.4 Cluster-Based Filtering

At this stage, the clustering results obtained previously are used to group songs based on the similarity of their audio characteristics. Each song in the dataset has been assigned a cluster label representing a specific music segment, such as high-energy tracks, acoustic songs, mellow-toned pieces, or instrumental-dominant tracks.

The utilization of these clusters aims to narrow the recommendation search space, where the recommendation process is no longer performed across the entire dataset but is instead focused on songs within the same cluster or those possessing similar characteristics. Through this approach, the system is able to reduce computational complexity and enhance the relevance of the recommended song candidates.

The results of the *Cluster-Based Filtering* implementation demonstrate that the song candidate selection process becomes more structured and contextual. The recommended songs exhibit audio characteristic similarities that are more consistent with user preferences, making this approach an effective initial stage for the implementation of the Deep Collaborative Filtering model.

3.5 Deep Collaborative Filtering (NeuMF)

In the *Deep Collaborative Filtering* (DCF) stage, the *Neural Matrix Factorization* (NeuMF) model is trained using user interaction data that has been filtered based on previous clustering results. The training process shows a consistent decrease in training loss until the final epoch, indicating that the model is able to learn user preference patterns stably. The model evaluation results yielded a *Hit Ratio@10 (HR@10)* of 0.1110 and a *Normalized Discounted Cumulative Gain@10 (NDCG@10)* of 0.0507. These values demonstrate that the model is capable of recommending relevant songs within the top recommendation list and organizing the recommendation order in accordance with user preferences.

The recommendation results generated for a single user show that the system is able to present songs from various genres, such as *Alternative Rock, Folk, Metal, Pop,* and *Rap*, while remaining within the relevant cluster space. This finding indicates that the integration of *Cluster-Based Filtering* with *Deep Collaborative Filtering* (NeuMF) can enhance recommendation relevance while simultaneously narrowing the *search space*, allowing the system to operate more effectively and adaptively. The recommendation results produced by the *NeuMF* model can be seen in Table 3.

Table 3. Song recommendation result using Deep Collaborative Filtering (NeuMF)

No	Judul Lagu	Artis	Tahun	Genre	Cluster
1	*Dead Leaves and the Dirty Ground*	The White Stripes	2001	Alternative Rock	0
2	*The Sinking of the Reuben James*	Cisco Houston	1962	Folk	3
3	*Jump Hi*	LION BABE	2014	Funk	0
4	*Chop Suey!*	System Of A Down	2001	Metal	0
5	*All That She Wants*	Ace of Base	1994	Pop	1
6	*Fast Car*	Tracy Chapman	1988	Pop	1
7	*Land: Horses / Land of a Thousand Dances / La…*	Patti Smith	1975	Punk	4
8	*On To The Next One*	JAY-Z	2009	Rap	4
9	*Freedom*	Pharrell Williams	2015	Rap	0
10	*Wishing For A Hero (feat. BJ The Chicago Kid)*	Polo G	2020	Rap	2

3.6 Evaluation and Analysis

In the evaluation and analysis stage, the performance of the Deep Collaborative Filtering (NeuMF) model is assessed to measure the quality of the generated recommendations. The evaluation is conducted using *Hit Ratio@10 (HR@10)* dan *Normalized Discounted Cumulative Gain@10 (NDCG@10)* metrics. Based on the testing results, the model achieved an HR@10 value of 0.1110 and an NDCG@10 value of 0.0507, indicating that the system is capable of recommending relevant songs within the user's top recommendation list. The HR@10 value signifies that relevant items successfully appeared within the top ten recommendations, while the NDCG@10 value shows that the generated recommendation order aligns well with the user's level of relevance. These results demonstrate that the integration of

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2609−2617
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9181

Cluster-Based Filtering with NeuMF can enhance the effectiveness of the recommendation system by narrowing the search space and delivering more relevant and targeted recommendations.

3.7 Discussion

The results of this study indicate that the integration of K-Means Clustering and Deep Collaborative Filtering can improve the effectiveness of music recommendation systems. The obtained evaluation results, with a Hit Ratio@10 of 0.1110 and an NDCG@10 of 0.0507, show that the proposed model is capable of recommending relevant songs within the top recommendation list. Compared with previous studies that primarily utilized clustering techniques only for music segmentation or descriptive analysis, this research extends the role of clustering by using cluster information as a filtering mechanism before the recommendation process, which helps reduce the search space of candidate songs. Furthermore, previous studies on Neural Collaborative Filtering have shown strong capability in capturing complex user–item interactions, but many of them rely solely on interaction data without incorporating item grouping strategies. By integrating clustering results with the NeuMF model, this study demonstrates that combining unsupervised learning with deep collaborative filtering can provide a more structured recommendation process while improving recommendation efficiency and personalization for large-scale music streaming platforms such as Spotify.

4. CONCLUSION

This research discusses the implementation of a *Hybrid Recommendation System* that integrates *K-Means Clustering* and *Deep Collaborative Filtering* based on *Neural Matrix Factorization* (NeuMF) to enhance the quality of personalized music recommendations on streaming platforms like Spotify. The results indicate that the clustering process using K-Means is capable of grouping more than 15,000 songs into five main clusters based on audio characteristic similarities, thereby forming clear music segmentations that are not concentrated in a single group. The utilization of these clustering results as a basis for *Cluster-based Filtering* is proven to narrow the *search space* for song candidates before being processed by the recommendation model, thus increasing system efficiency and relevancy. Furthermore, the implementation of the *Deep Collaborative Filtering* (NeuMF) model shows reasonably good performance based on evaluation results using *Hit Ratio@10* and *NDCG@10* metrics, with values of 0.1110 and 0.0507, respectively. These values indicate that the system is able to recommend relevant songs and arrange the recommendation order adequately. The integration of this unsupervised learning approach with a deep learning-based recommender system shows significant potential in addressing common issues in conventional recommendation systems, such as personalization limitations and the *cold start* problem.This study contributes to the development of music recommendation systems by demonstrating that the integration of clustering-based item grouping with deep collaborative filtering can improve recommendation relevance while simultaneously reducing the computational search space in large-scale music datasets. Nevertheless, this study still has limitations, particularly in the use of synthetic interaction data which may not fully represent real user behavior, as well as limitations in the evaluation metrics employed. Therefore, future research is expected to utilize direct user interaction data, increase the variety of evaluation metrics, and explore more complex clustering methods or recommendation model architectures to improve the performance and reliability of music recommendation systems.

REFERENCES

[1] P. B. Chopade, “Music Recommendation System on Spotify Using Deep Learning,” *Int. J. Sci. Res. Eng. Manag.*, vol. 09, no. 06, pp. 1–9, 2025, doi: 10.55041/ijsrem49675.
[2] K. R. Putra and M. A. Rachman, “Perbandingan Metode Content-based, Collaborative dan Hybrid Filtering pada Sistem Rekomendasi Lagu,” *MIND J.*, vol. 9, no. 2, pp. 179–193, 2024, doi: 10.26760/mindjournal.v9i2.179-193.
[3] M. A. Munajad, A. Ridwan, and T. G. Pratama, “Pengembangan Sistem Rekomendasi Musik dengan K-Means dan KNN Berbasis Cosine Similarity,” *Sainteks*, vol. 22, no. 2, pp. 153–165, 2025, doi: 10.30595/sainteks.v22i2.27815.
[4] D. Roy and M. Dutta, "A Systematic Review and Research Perspective on Recommender Systems," Journal of Big Data, vol. 9, no. 1, 2022, doi: 10.1186/s40537-022-00592-5
[5] N. K. Ayyiyah, R. Kusumaningrum, and R. Rismiyati, “Film Recommender System Menggunakan Metode Neural Collaborative Filtering,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 3, pp. 699–708, 2023, doi: 10.25126/jtiik.2023106616.
[6] B. Hakim, F. J. Kaunang, C. Susanto, J. Salim, and R. Indradjaja, “Implementasi Machine Learning Dalam Pengelompokan Musik Menggunakan Algoritma K-Means Clustering,” *IDEALIS Indonesian Journal of Information Systems.*, vol. 8, no. 1, pp. 74–83, 2025, doi: 10.36080/idealis.v8i1.3357.
[7] M. I. Firmansyah, R. S. Rohman, and Marsusanti, "Penerapan Algoritma Klastering K-Means untuk Fitur Atribut pada layanan streaming Spotify," *Indonesian Journal Computer Science*, vol. 2, no. 2, 2023, doi: 10.31294/ijcs.v2i2.2465
[8] K. B. Dharmasena and C. Pramartha, “Segmentasi Pengguna Spotify Berdasarkan Preferensi Musik dengan Algoritma K-Means Clustering,” *Jnatia*, vol. 3, no. 1, pp. 37–42, 2024. doi: 10.24843/JNATIA.2024.v03.i01.p05.
[9] M. D. Noverta Effendi*1, Witrihan Ramadhani2, Fitri Farida3, “Jurnal Computer Science and Information Technology (CoSciTech) things,” *J. Comput. Sci. Inf. Technol.*, vol. 5, no. 2, pp. 358–366, 2024, doi: 10.37859/conscitech.v5i2.5600.
[10] Y. Lase and E. Panggabean, “Implementasi Metode K-Means Clustering Dalam Sistem Pemilihan Jurusan Di SMK Swasta Harapan Baru,” *J. Teknol. dan Ilmu Komput. Prima*, vol. 2, no. 2, p. 43, 2019, doi: 10.34012/jutikomp.v2i2.723.
[11] M. Cui, “on the Elbow Method,” *Clausius Sci. Press*, vol. 1, no. 1, pp. 5–8, 2020, doi: 10.23977/accaf.2020.010102.
[12] D. M. Saputra, D. Saputra, and L. D. Oswari, “Effect of Distance Metrics in Determining K-Value in K-Means Clustering

Building of Informatics, Technology and Science (BITS)
Volume 7, No 4, March 2026 Page: 2609−2617
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i4.9181

Using Elbow and Silhouette Method," Siconian 2019: *Sriwijaya International Conference on Information Techonoly and Its Application*, vol. 172, pp. 341–346, 2020, doi: 10.2991/aisr.k.200424.051.

[13] B. A. Firdaus, D. E. Ratnawati, and B. T. Hanggara, "Klusterisasi Popularitas Artist pada Playlist Today's Top Hits Menggunakan Metode K-Means dengan Integrasi Spotify Web API dan Teknologi Amazon SageMaker," *J. Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 1, pp. 373–380, 2021, doi: 10.25126/j-ptiik.202151373.

[14] M. A. Munajad, A. Ridwan, and T. G. Pratama, "Pengembangan Sistem Rekomendasi Musik dengan K-Means dan K-Nearest Neighbors berbasis Cosine Similarity," Sainteks: *jurnal Sains dan Teknologi*, vol. 22. no. 2, 2024, doi: 10.30595/sainteks.v22i2.27815.

[15] D. D. Satrio, F. A. Akbar, and M. M. Al Haromainy, "Pengembangan Bot Discord Sebagai Pemutar dan Rekomendasi Musik Menggunakan Metode K-Means," *Jutisi: Jurnal Ilmu Teknologi Informormasi dan Sistem Informasi*, vol. 13, no. 1, p. 95, 2024, doi: 10.35889/jutisi.v13i1.1681.

[16] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T. S. Chua, "Neural Collaborative Filtering," Proceedings of the 26th International Conference on World Wide Web, January, 2021, doi: 10.1145/3038912.3052569.

[17] Y. T. Bau, Reza, and K. C. Lee, "Developing and Comparing Machine Learning Algorithm for Music Recommendation," JOIV: International Journal on Informatics Visualization, vol. 8, no. 3-2, 2024, doi: 10.62527/joiv.8.3-2.2947.

[18] Indah Lestari, "Analisis Clustering Lagu Berdasarkan Fitur Audio Pada Spotify Untuk Identifikasi Genre Musik Menggunakan Algoritma K-Means," vol. 32, no. 3, 2023, pp. 167–186, 2021.

[19] M. M. Raharjo and F. Arifin, "Machine Learning System Implementation of Education Podcast Recommendation on Spotify Application Using Content-Based Filtering and TF-IDF," ELINVO: *Electronics, INformatics, and Vocational Education,* vol. 8, no. 2, 2023, doi: 10.21831/elinvo.v8i2.58014.

[20] R. C. de Araujo, V. M. S. Santos, J. F. L. de Oliveira, and A. M. A. Maciel, "A Hybrid Music Recommendation System Based on K-Means Clustering and Multilayer Perceptron," *Int. Conf. Enterp. Inf. Syst. ICEIS - Proc.*, vol. 1, no. Iceis, pp. 335–342, 2025, doi: 10.5220/0013436700003929.