

ABSTRAK

Sistem rekomendasi lagu berbasis *Collaborative Filtering* menghadapi keterbatasan fundamental berupa *cold-start problem* ketika berhadapan dengan lagu baru atau pengguna baru yang belum memiliki riwayat interaksi. Penelitian ini mengimplementasikan pendekatan *Content-Based Filtering* (CBF) menggunakan algoritma *K-Means Clustering* berbasis ekstraksi fitur *audio* untuk mengatasi permasalahan tersebut. *Dataset* yang digunakan terdiri dari 3.200 *file* lagu dalam format *mp3*, *wav*, *aac*, dan *m4a* yang dikumpulkan dari penyimpanan lokal. Fitur *audio* diekstraksi menggunakan library *Librosa*, menghasilkan representasi numerik berupa *Mel-Frequency Cepstral Coefficients* (MFCC), *Spectral Centroid*, *Zero Crossing Rate* (ZCR), *RMS Energy*, *Spectral Rolloff*, *Chroma STFT*, *Tempo*, dan *Beat Count*. Setelah normalisasi *Z-Score* untuk menyeragamkan skala fitur, algoritma *K-Means Clustering* diterapkan dan dievaluasi menggunakan *Within-Cluster Sum of Squares* (WCSS) dan *Calinski-Harabasz* (CH) Index, menggunakan partisi optimal pada $k=6$ dalam *dataset*, dengan nilai *WCSS* **95.616,62** dan *CH Index* **85,92**. Sistem rekomendasi menghasilkan *output Top-N* menggunakan dua metrik jarak yaitu *Euclidean Distance* dan *Cosine Similarity*, yang dievaluasi menggunakan *Overlap Coefficient*, *Jaccard Index*, dan *Intra-List Similarity* (ILS). Hasil evaluasi menunjukkan konvergensi mutlak antara kedua metode pada **Top-3** dengan nilai *Jaccard Index* **1,0**, sementara rata-rata skor *ILS* stabil di rentang **0,5** yang mengindikasikan keseimbangan optimal antara relevansi dan diversitas rekomendasi. Sistem diimplementasikan dalam aplikasi *web* berbasis *Flask* dengan visualisasi *Diagram Venn* dan *Heatmap* sebagai pendukung evaluasi interaktif.

Kata Kunci: Sistem Rekomendasi Lagu, *K-Means*, *Machine Learning*, Ekstraksi *Audio*, *Librosa*, *Mel-Frequency Cepstral Coefficients*, *Intra-List Similarity*, *Jaccard Similarity*

ABSTRACT

Music recommender systems based on Collaborative Filtering face a fundamental limitation known as the cold-start problem when dealing with new songs or new users who have no prior interaction history. This study implements a Content-Based Filtering (CBF) approach using the K-Means Clustering algorithm based on audio feature extraction to address this limitation. The dataset consists of 3,200 audio files in mp3, wav, aac, and m4a formats collected from local storage. Audio features were extracted using the Librosa library, producing numerical representations including Mel-Frequency Cepstral Coefficients (MFCC), Spectral Centroid, Zero Crossing Rate (ZCR), RMS Energy, Spectral Rolloff, Chroma STFT, Tempo, and Beat Count. Following Z-Score normalization to standardize feature scales, the K-Means Clustering algorithm was applied and evaluated using Within-Cluster Sum of Squares (WCSS) and the Calinski-Harabasz (CH) Index, yielding an optimal partition at $k=6$ in the dataset, with a WCSS value of **95,616.62** and a CH Index of **85.92**. The recommender system produces Top-N outputs using two distance metrics, namely Euclidean Distance and Cosine Similarity, evaluated through Overlap Coefficient, Jaccard Similarity, and Intra-List Similarity (ILS). Evaluation results demonstrate absolute convergence between both methods at **Top-3** with a Jaccard Similarity of **1.0**, while the average ILS score remains stable around **0.5**, indicating an optimal balance between recommender relevance and diversity. The system is implemented as a Flask-based web application featuring Venn Diagram and Heatmap visualizations to support interactive evaluation for users.

Keywords: Music Recommender System, K-Means, Content-Based Filtering, Audio Feature Extraction, Librosa, Mel-Frequency Cepstral Coefficients, Intra-List Similarity, Jaccard Similarity