

**PENDEKATAN HYBRID UNTUK REKOMENDASI MUSIK  
BERDASARKAN KEMIRIPAN AUDIO MENGGUNAKAN  
AUTOENCODER DAN LIGHTGBM**

**LAPORAN NON-REGULER**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



Disusun oleh :

**WINDA ARDELIA ARISTAWIDYA**

**22.11.4852**

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2025**

**PENDEKATAN HYBRID UNTUK REKOMENDASI MUSIK  
BERDASARKAN KEMIRIPAN AUDIO MENGGUNAKAN  
AUTOENCODER DAN LIGHGBM**

**LAPORAN NON-REGULER**

Diajukan untuk memenuhi salah satu syarat mencapai derajat  
Sarjana Program Studi Informatika



Disusun oleh :

**WINDA ARDELIA ARISTAWIDYA**

**22.11.4852**

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2025**

HALAMAN PERSETUJUAN

JALUR NON-REGULER

**PENDEKATAN HYBRID UNTUK REKOMENDASI MUSIK  
BERDASARKAN KEMIRIPAN AUDIO MENGGUNAKAN  
AUTOENCODER DAN LIGHTGBM**

yang disusun dan diajukan oleh

**Winda Ardella Aristawidya**

**22.11.4852**

telah disetujui oleh Dosen Pembimbing  
pada tanggal 15 Desember 2025

**Dosen Pembimbing,**



**Majid Rahardi, S.Kom., M.Eng**  
**NIK. 190302393**

HALAMAN PENGESAHAN

JALUR NON-REGULER

**PENDEKATAN HYBRID UNTUK REKOMENDASI MUSIK BERDASARKAN  
KEMIRIPAN AUDIO MENGGUNAAN AUTOENCODER DAN LIGHTGBM**

yang disusun dan diajukan oleh  
**Winda Ardella Aristawidya**  
22.11.4852

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 15 Desember 2025

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Arifvanto Hadinegoro, S.Kom., M.T.**  
NIK. 190302289

**Dr. Hartatik, S.T., M.Cs.**  
NIK. 190302232

**Majid Rahardi, S.Kom., M.Eng.**  
NIK. 190302393

Laporan ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 15 Desember 2025

**DEKAN FAKULTAS ILMU KOMPUTER**



**Prof. Dr. Kusriani, M.Kom.**  
NIK. 190302106

## HALAMAN PERNYATAAN KEASLIAN KARYA

Yang bertandatangan di bawah ini,

Nama mahasiswa : Winda Ardelia Aristawidya

NIM : 22.11.4852

Menyatakan bahwa laporan dengan judul berikut:

Pendekatan hybrid untuk rekomendasi musik berdasarkan kemiripan audio menggunakan autoencoder dan lightgbm.

Dosen Pembimbing : Majid Rahardi, S.Kom., M.Eng.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan kegiatan SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidak-benaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 15 Desember 2025

Yang Menyatakan,

Winda Ardelia Aristawidya

## HALAMAN PERSEMBAHAN

Pertama-tama saya panjatkan syukur kepada Allah SWT karena atas berkah an ridha nya saya bisa menyelesaikan karya ilmiah ini dengan baik, dan tidak lupa karya penelitian ini saya persembahkan kepada :

1. Kedua orang tua saya yang senantiasa memfasilitasi saya dengan kesempatan untuk menempuh jenjang pendidikan, serta memberikan tenaga, materi, dan pikiran, sehingga saya dapat mencapai tahap ini. Semoga Allah SWT senantiasa memberkahi kehidupan mereka.
2. Dosen Pembimbing yang sudah membantu saya dengan baik, sabar dan bijaksana dalam proses pengerjaan karya ilmiah ini. Terima kasih atas dedikasi yang sudah bapak berikan. Semoga segala kebaikan mendapatkan balasan dari Allah SWT.
3. Teman-teman dan juga sahabat yang selalu memberikan dukungan dalam kondisi apapun.
4. Universitas Amikom Yogyakarta yang menjadi tempat saya belajar, berkembang, dan memperoleh berbagai pengalaman berharga selama masa perkuliahan.
5. Xdinary Heroes, DAY6 dan Eaj yang menjadi awal inspirasi saya dalam membuat karya ilmiah ini. Berkat karya-karya yang mereka hasilkan, saya dapat menemukan motivasi, semangat, serta perspektif baru yang akhirnya mendorong saya untuk menyelesaikan karya ilmiah ini dengan sepenuh hati.

## KATA PENGANTAR

Puji syukur kehadiran Allah SWT, berkat rahmat dan ridha nya yang telah membantu dalam menyelesaikan karya ilmiah dengan dengan judul "PENDEKATAN HYBRID UNTUK REKOMENDASI MUSIK BERDASARKAN KEMIRIPAN AUDIO MENGGUNAKAN AUTOENCODER DAN LIGHGBM". Penelitian ini disusun untuk menyelesaikan kewajiban dalam memenuhi persyaratan kelulusan program studi Informatika fakultas ilmu komputer Univeristas Amikom Yogyakarta. Penyusunan karya ilmiah ini dapat terselesaikan berkat doa, dukungan, bantuan, serta bimbingan dari berbagai pihak. Untuk itu, penulis menyampaikan apresiasi dan terima kasih yang sebesar-besarnya kepada:

1. Keluarga yang sudah mendukung secara materil dan moral dalam penyelesaian karya ilmiah ini.
2. Majid Rahardi, S.Kom., M.Eng. selaku dosen pembimbing yang telah memberikan bimbingan, masukan, serta dorongan yang berarti selama seluruh proses penyusunan penelitian ini. Dukungan beliau menjadi bagian penting dalam terselesaikannya karya ilmiah ini.
3. Teman-teman dan sahabat yang lapang tangan membantu penulis dalam penyusunan karya ilmiah ini.
4. Setiap pihak yang telah berbagi ide, masukan, maupun bantuan dalam bentuk apa pun, sehingga turut mendukung terselesaikannya penyusunan karya ilmiah ini.

Yogyakarta, 26 November 2025

Penulis

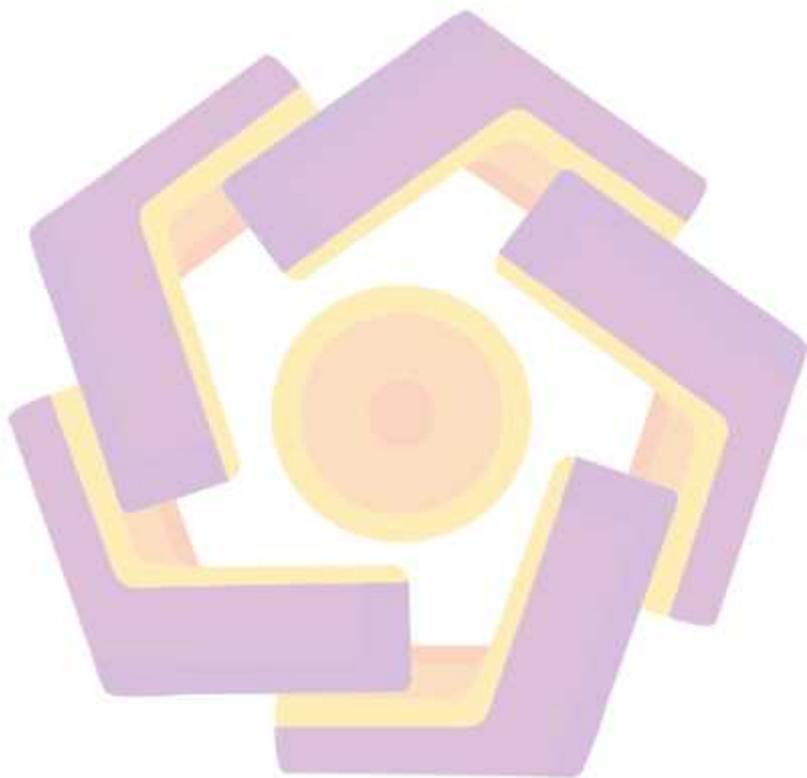
## DAFTAR ISI

|  |     |
|--|-----|
| HALAMAN JUDUL.....                         | i   |
| HALAMAN PERSETUJUAN.....                   | ii  |
| HALAMAN PENGESAHAN.....                    | iii |
| HALAMAN PERNYATAAN KEASLIAN KARYA.....     | iv  |
| HALAMAN PERSEMBAHAN.....                   | v   |
| KATA PENGANTAR.....                        | vi  |
| DAFTAR ISI.....                            | vii |
| DAFTAR TABEL.....                          | ix  |
| DAFTAR GAMBAR.....                         | x   |
| DAFTAR LAMBANG DAN SINGKATAN.....          | xi  |
| DAFTAR ISTILAH.....                        | xii |
| INTISARI.....                              | xiv |
| <i>ABSTRACT</i> .....                      | xv  |
| <b>BAB I PENDAHULUAN</b> .....             | 1   |
| 1.1. Gambaran Umum.....                    | 1   |
| 1.2. Rumusan Masalah.....                  | 3   |
| 1.3. Batasan Masalah.....                  | 3   |
| 1.4. Tujuan.....                           | 4   |
| <b>BAB II TINJAUAN PUSTAKA</b> .....       | 5   |
| 1.1. Studi Literatur.....                  | 5   |
| 1.2. Landasan Teori.....                   | 7   |
| 2.2.1 Sistem Rekomendasi Musik.....        | 7   |
| 2.2.2 Fitur Audio.....                     | 7   |
| 2.2.3 Data Preparation.....                | 8   |
| 2.2.4 Principal Component Analysis.....    | 9   |
| 2.2.5 Autoencoder.....                     | 10  |
| 2.2.6 K-Nearest Neighbors.....             | 10  |
| 2.2.7 Cosine Similarity.....               | 10  |
| 2.2.8 Light Gradient Boosting Machine..... | 11  |
| 2.2.9. Confusion Matrix.....               | 11  |
| 2.2.10. 5-Fold Cross-Validation.....       | 11  |
| <b>BAB III METODE PENELITIAN</b> .....     | 13  |
| 3.1. Metode.....                           | 13  |

|  |  |           |
|--|--|-----------|
| 3.1.1.                                   | Data Collection.....                     | 14        |
| 3.1.2.                                   | Data Preparation.....                    | 14        |
| 3.1.3.                                   | Rekayasa Fitur.....                      | 15        |
| 3.1.5.                                   | Evaluasi Model.....                      | 16        |
| 3.1.6.                                   | Implementasi dan Interaksi Pengguna..... | 17        |
| <b>BAB IV PEMBAHASAN.....</b>            |  | <b>19</b> |
| 4.1.                                     | Data Collection.....                     | 19        |
| 4.2.                                     | Data Preparation.....                    | 20        |
| 4.3.                                     | Rekayasa Fitur.....                      | 22        |
| 4.3.1.                                   | Principal Component Analysis.....        | 22        |
| 4.3.2.                                   | Autoencoder.....                         | 23        |
| 4.4.                                     | Pengembangan Model.....                  | 25        |
| 4.4.1.                                   | K-Nearest Neighbor.....                  | 25        |
| 4.4.2.                                   | Light Gradient Boost Machine.....        | 26        |
| 4.5.                                     | Evaluasi Model.....                      | 26        |
| 4.5.1.                                   | Confusion Matrix.....                    | 27        |
| 4.5.2.                                   | 5-fold Cross-Validation.....             | 28        |
| 4.6.                                     | Implementasi dan Interaksi Pengguna..... | 28        |
| <b>BAB V KESIMPULAN.....</b>             |  | <b>30</b> |
| 4.1.                                     | Kesimpulan.....                          | 30        |
| 4.2.                                     | Saran.....                               | 30        |
| <b>REFERENSI.....</b>                    |  | <b>31</b> |
| <b>CURICULUM VITAE.....</b>              |  | <b>33</b> |
| <b>LAMPIRAN DAN BUKTI PENDUKUNG.....</b> |  | <b>35</b> |
| a.                                       | Letter of Acceptance.....                | 35        |
| b.                                       | Lembar Review.....                       | 36        |
| c.                                       | Bukti Terbit/Terindex.....               | 39        |
| d.                                       | Bukti pembayaran.....                    | 40        |

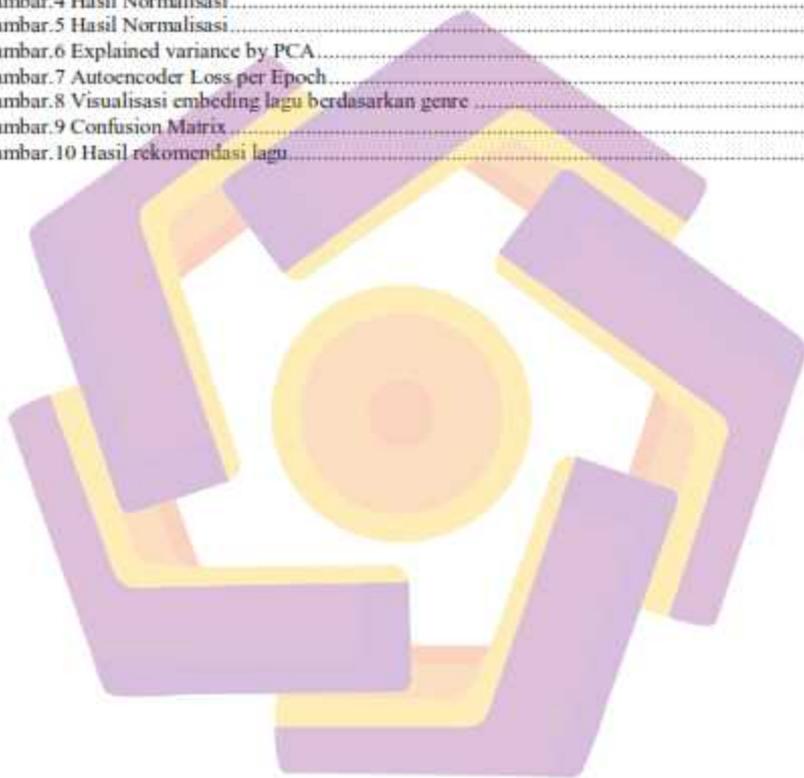
## DAFTAR TABEL

|  |    |
|--|----|
| Table 3.1.1 Confusion Matrix .....                       | 17 |
| Table 4.1.1 Isi dataset fitur audio .....                | 19 |
| Table 4.5.1 Hasil evaluasi 5-fold cross-validation ..... | 28 |



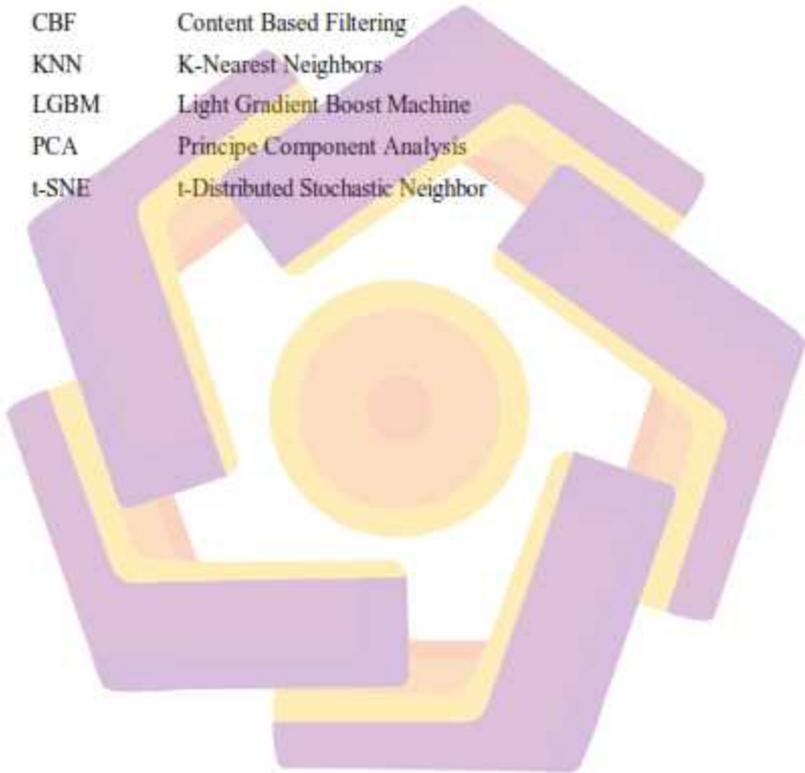
## DAFTAR GAMBAR

|   |    |
|---|----|
| Gambar.1 Alur Penelitian .....                              | 13 |
| Gambar.2 Jumlah Dataset .....                               | 20 |
| Gambar.3 Kolom Dataset .....                                | 21 |
| Gambar.4 Hasil Normalisasi .....                            | 21 |
| Gambar.5 Hasil Normalisasi .....                            | 22 |
| Gambar.6 Explained variance by PCA .....                    | 22 |
| Gambar.7 Autoencoder Loss per Epoch .....                   | 24 |
| Gambar.8 Visualisasi embedding lagu berdasarkan genre ..... | 25 |
| Gambar.9 Confusion Matrix .....                             | 27 |
| Gambar.10 Hasil rekomendasi lagu .....                      | 29 |



## DAFTAR LAMBANG DAN SINGKATAN

|              |   |
|--------------|---|
| $Z_{scaled}$ | Nilai hasil Normalisasi                   |
| $X$          | Nilai asli dari fitur sebelum normalisasi |
| $\mu$        | Mean ( $\mu$ )                            |
| $\sigma$     | Standar Deviasi ( $\sigma$ )              |
| CBF          | Content Based Filtering                   |
| KNN          | K-Nearest Neighbors                       |
| LGBM         | Light Gradient Boost Machine              |
| PCA          | Principle Component Analysis              |
| t-SNE        | t-Distributed Stochastic Neighbor         |

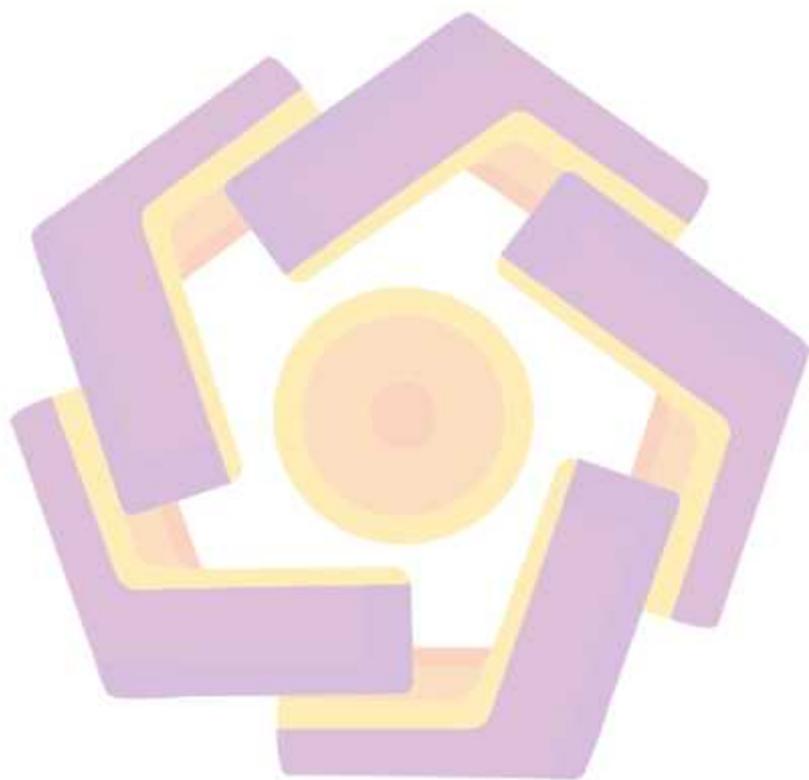


## DAFTAR ISTILAH

|                          |   |
|--------------------------|---|
| Accuracy                 | Ukuran seberapa tepat model dalam memprediksi kelas atau rekomendasi.   |
| Autoencoder              | Model neural network untuk mengekstraksi representasi ringkas (embedding) dari data.                                |
| Content-Based Filtering  | Metode rekomendasi yang mencari item mirip berdasar fitur konten.   |
| Cosine Similarity        | Metode untuk mengukur tingkat kemiripan dua vektor dengan menghitung sudut di antara keduanya.                      |
| Data Preparation         | Tahap awal pengolahan data untuk membuat data lebih bersih, terstruktur, dan siap digunakan dalam proses pemodelan. |
| Dimensionality Reduction | Mengurangi jumlah fitur agar model lebih efisien.   |
| Embedding                | Representasi numerik berdimensi kecil yang menangkap karakteristik penting dari data.                               |
| Feature Engineering      | Proses membuat atau memodifikasi fitur agar lebih informatif untuk model.   |
| Feature Selection        | Pemilihan fitur yang paling relevan untuk meningkatkan performa model.  |
| KNN                      | Algoritma yang mencari data paling mirip berdasarkan jarak antar vektor.  |
| LightGBM Ranker          | Algoritma learning-to-rank berbasis gradient boosting untuk mereranking hasil rekomendasi.                          |
| Model Evaluation         | Proses menilai performa model menggunakan metrik tertentu.  |
| nDCG                     | Metrik evaluasi untuk sistem rekomendasi yang mengukur kualitas urutan rekomendasi.                                 |
| Normalisasi              | Proses menyamakan skala fitur agar model lebih stabil.  |
| PCA                      | Teknik reduksi dimensi untuk menangkap variasi terbesar dalam data.   |
| Re-ranking               | Tahap mengurutkan kembali hasil rekomendasi awal menggunakan model tambahan.  |
| StandardScaler           | Teknik normalisasi yang mengubah data menjadi distribusi dengan mean 0 dan standar deviasi 1.                       |
| Sistem Rekomendasi       | Sistem yang memberikan saran item berdasarkan preferensi atau pola pengguna.  |

Training  
t-SNE

Proses model belajar dari data untuk mengenali pola.  
Teknik visualisasi untuk melihat sebaran embedding dalam dua dimensi.



## INTISARI

Sistem rekomendasi musik berperan penting dalam membantu pengguna menelusuri koleksi musik yang sangat luas dengan memberikan saran lagu yang sesuai dengan preferensi mereka. Namun, metode konvensional sering kali kurang memperhatikan kedalaman konten audio, sehingga menyebabkan keterbatasan dalam aspek personalisasi dan akurasi rekomendasi. Penelitian ini mengusulkan pendekatan hybrid yang memanfaatkan Principal Component Analysis (PCA) dan Autoencoder untuk mengekstraksi audio embedding. Hasil embedding tersebut kemudian diproses menggunakan metode K-Nearest Neighbors (KNN) untuk menemukan lagu dengan kemiripan tinggi, yang selanjutnya dilakukan tahap reranking menggunakan Light Gradient Boosting Machine (LightGBM) berdasarkan relevansi yang diprediksi. Sistem yang dikembangkan menunjukkan hasil yang kuat dengan tingkat akurasi sebesar 98%, precision 0.96, recall 0.96, dan F1-score 0.96 untuk kelas Similar, serta precision dan recall 0.99 untuk kelas Not Similar. Proses cross-validation juga mengonfirmasi keandalan model dengan rata-rata akurasi 97.99%, precision 0.9577, recall 0.9624, dan F1-score 0.9600, disertai deviasi standar yang rendah. Hasil tersebut menunjukkan bahwa kombinasi antara fitur audio yang diekstraksi secara mendalam dengan metode machine learning berbasis ranking dapat meningkatkan kualitas rekomendasi musik secara signifikan. Untuk pengembangan di masa mendatang, sistem ini dapat ditingkatkan dengan menambahkan data metadata serta visualisasi berbasis genre agar hasil yang diperoleh menjadi lebih beragam dan mudah diinterpretasikan.

Kata kunci: Sistem Rekomendasi Musik, Fitur Audio, Autoencoder, PCA, LightGBM.

## ABSTRACT

*Music recommendation systems help users navigate large music collections by suggesting songs aligned with their preferences. However, conventional methods often overlook the depth of audio content, limiting personalization and accuracy. This study proposes a hybrid approach that uses PCA and Autoencoder to extract audio embeddings. These embeddings are processed using K-Nearest Neighbors to find similar tracks, followed by a reranking step with LightGBM based on predicted relevance. The system achieved strong results: 98% accuracy, 0.96 precision, 0.96 recall, and 0.96 F1-score for the Similar class, with 0.99 precision and recall for Not Similar. Cross-validation confirmed model robustness, with an average accuracy of 97.99%, precision of 0.9577, recall of 0.9624, and F1-score of 0.9600, all with low standard deviations. These outcomes show that combining deep audio features with machine learning ranking enhances recommendation quality. Future improvements may involve incorporating metadata and genre-based visualizations for more diverse and interpretable results.*

**Keyword:** *Music recommendation System, Audio feature, Autoencoder, PCA, LightGBM*