



SKRIPSI

**PERSONALISASI REKOMENDASI MUSIK UNTUK KEGIATAN BERKENDARA,
BELAJAR, DAN MAKAN MENGGUNAKAN HYBRID MATRIX
FACTORIZATION BERBASIS ALS DAN HDBSCAN PADA MILLION SONG
DATASET**

BADAR KATAMSI

NPM 21081010186

DOSEN PEMBIMBING

Fawwaz Ali Akbar, S.Kom., M.Kom.

Achmad Junaidi, S.Kom., M.Kom.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
2025**



SKRIPSI

**PERSONALISASI REKOMENDASI MUSIK UNTUK KEGIATAN BERKENDARA,
BELAJAR, DAN MAKAN MENGGUNAKAN HYBRID MATRIX
FACTORIZATION BERBASIS ALS DAN HDBSCAN PADA MILLION SONG
DATASET**

BADAR KATAMSI

NPM 21081010186

DOSEN PEMBIMBING

Fawwaz Ali Akbar, S.Kom., M.Kom.

Achmad Junaidi, S.Kom., M.Kom.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
2025**

Halaman ini sengaja dikoso

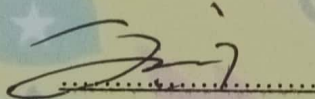
LEMBAR PENGESAHAN

PERSONALISASI REKOMENDASI MUSIK UNTUK KEGIATAN BERKENDARA, BELAJAR, DAN MAKAN MENGGUNAKAN HYBRID MATRIX FACTORIZATION BERBASIS ALS DAN HDBSCAN PADA MILLION SONG DATASET

Oleh :
BADAR KATAMSI
NPM. 21081010186

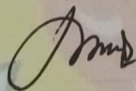
Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 1 September 2025

Fawwaz Ali Akbar, S.Kom., M.Kom.
NIP. 19700619 202121 1 009



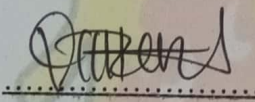
(Pembimbing I)

Achmad Junaidi, S.Kom., M.Kom.
NPT. 3 7811 04 0199 1



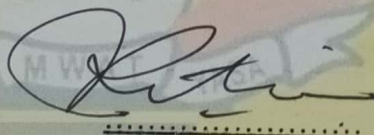
(Pembimbing II)

Henni Endah Wahanani, S.T, M.Kom.
NIP. 19780922 202121 2 005



(Penguji I)

Retno Mumpuni, S.Kom, M.Sc.
NPT. 172198 70 716054



(Penguji II)

Mengetahui,
Dekan Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP. 19681126 199403 2 001

Halaman ini sengaja dikosongkan

LEMBAR PERSETUJUAN

**PERSONALISASI REKOMENDASI MUSIK UNTUK KEGIATAN
BERKENDARA, BELAJAR, DAN MAKAN MENGGUNAKAN HYBRID
MATRIX FACTORIZATION BERBASIS ALS DAN HDBSCAN PADA
MILLION SONG DATASET**

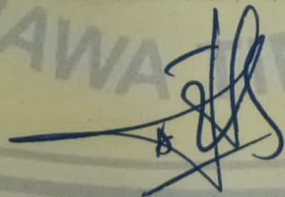
Oleh :

BADAR KATAMSI

NPM 21081010186

Menyetujui

Koordinator Program Studi Informatika
Fakultas Ilmu Komputer,



Fetty Tri Anggraeny, S.Kom., M.Kom.
NIP. 19820211 2021212 005

Halaman ini sengaja dikosongkan

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan dibawah ini :

Nama : Badar Katamsi
NPM : 21081010186
Program : Sarjana (S1)
Program Studi : Informatika
Fakultas : Fakultas Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Tugas Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila dikemudian hari ditemulan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.



Surabaya, 1 September 2025

nyataan,

NPM. 21081010186

Halaman ini sengaja dikosongkan

ABSTRAK

Nama Mahasiswa / NPM : Badar Katamsi / 21081010186

Judul Skripsi : Personalisasi Rekomendasi Musik untuk Kegiatan Berkendara, Belajar, dan Makan menggunakan Hybrid Matrix Factorization berbasis ALS dan HDBSCAN pada Million Song Dataset

Dosen Pembimbing : 1. Fawwaz Ali Akbar, S.Kom., M.Kom.
2. Achmad Junaidi, S.Kom., M.Kom.

Perkembangan platform streaming musik dengan katalog lebih dari 100 juta lagu, seperti Spotify, menimbulkan tantangan *information overload* yang membuat pengguna kesulitan menemukan musik sesuai preferensi, sementara masalah *cold start* pada pengguna baru juga membatasi akurasi sistem rekomendasi. Penelitian ini bertujuan mengembangkan sistem rekomendasi musik kontekstual yang mampu mengatasi kedua masalah tersebut dengan mengintegrasikan *Matrix Factorization* berbasis *Alternating Least Squares (ALS)* dan algoritma *HDBSCAN* sebagai metode klusterisasi, serta menambahkan filter berbasis fitur audio untuk menyesuaikan rekomendasi dengan konteks kegiatan pengguna, yaitu berkendara, belajar, dan makan. Dataset yang digunakan adalah *Million Song Dataset* dari Spotify & Last.fm dengan 50.683 lagu dan hampir satu juta riwayat dengar pengguna. Tahap penelitian meliputi pembersihan data, transformasi playcount menjadi rating eksplisit, ekstraksi fitur genre menggunakan TF-IDF, reduksi dimensi dengan UMAP, klusterisasi menggunakan HDBSCAN, hingga pelatihan model ALS pada setiap kluster. Hasil evaluasi menunjukkan nilai *Silhouette Score* sebesar 1,00 yang menandakan kluster optimal, sementara model ALS terbaik mencapai *Reconstruction Error* sebesar 0,000814, *RMSE* sebesar 0,1525, dan *MAE* sebesar 0,0630. Survei kepuasan pengguna juga memperlihatkan skor rata-rata 4,2 dari 5 dengan 80% lagu rekomendasi disimpan ke dalam playlist. Dibandingkan pendekatan tradisional berbasis ALS tunggal, metode hybrid yang diusulkan terbukti lebih efektif dalam mengatasi sparsity, *cold start*, serta meningkatkan relevansi rekomendasi berbasis *occasion*. Kesimpulannya, integrasi ALS dan HDBSCAN dengan filter fitur audio dapat menghasilkan sistem rekomendasi musik yang efisien, akurat, dan lebih personal, sehingga berkontribusi pada peningkatan kualitas pengalaman pengguna pada platform musik digital.

Kata kunci : Rekomendasi Musik, ALS, HDBSCAN, Sistem Rekomendasi, Clustering

Halaman ini sengaja dikosongkan

ABSTRACT

Student Name / NPM : Badar Katamsi / 21081010186

Thesis Title : Personalized Music Recommendations for Driving, Studying, and Eating Using Hybrid Matrix Factorization Based on ALS and HDBSCAN on a Million Song Dataset

Advisor : 1. Fawwaz Ali Akbar, S.Kom., M.Kom.
2. Achmad Junaidi, S.Kom., M.Kom.

ABSTRACT

The development of music streaming platforms with catalogs of more than 100 million songs, such as Spotify, poses the challenge of information overload, making it difficult for users to find music that suits their preferences, while the cold start problem for new users also limits the accuracy of the recommendation system. This study aims to develop a contextual music recommendation system capable of addressing both issues by integrating Matrix Factorization based on Alternating Least Squares (ALS) and the HDBSCAN algorithm as clustering methods, along with audio feature-based filters to tailor recommendations to the user's activity context, namely driving, studying, and eating. The dataset used is the Million Song Dataset from Spotify & Last.fm, comprising 50,683 songs and nearly one million user listening histories. The research stages include data cleaning, transforming playcount into explicit ratings, extracting genre features using TF-IDF, dimension reduction with UMAP, clustering using HDBSCAN, and training the ALS model on each cluster. Evaluation results showed a Silhouette Score of 1.00, indicating optimal clustering, while the best ALS model achieved a Reconstruction Error of 0.000814, RMSE of 0.1525, and MAE of 0.0630. A user satisfaction survey also revealed an average score of 4.2 out of 5, with 80% of recommended songs saved to playlists. Compared to the traditional single ALS-based approach, the proposed hybrid method proved more effective in addressing sparsity, cold start, and improving the relevance of occasion-based recommendations. In conclusion, the integration of ALS and HDBSCAN with audio feature filters can produce an efficient, accurate, and more personalized music recommendation system, thereby contributing to improved user experience quality on digital music platforms.

Keywords: Music Recommendation, ALS, HDBSCAN, System Recommendation, Clustering

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul **“Personalisasi Rekomendasi Musik untuk Kegiatan Berkendara, Belajar, dan Makan menggunakan Hybrid Matrix Factorization berbasis ALS dan HDBSCAN pada Million Song Dataset”** dapat terselesaikan dengan baik. Penelitian ini merupakan salah satu syarat untuk menyelesaikan studi di Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Dalam penyusunan skripsi ini, penulis mendapat banyak bantuan, dukungan, dan bimbingan dari berbagai pihak. Oleh karena itu, dengan penuh rasa hormat dan terima kasih, penulis mengucapkan terima kasih sebesar – besarnya kepada:

1. Kedua orang tua tersayang, yang senantiasa memberikan dukungan, doa, serta kasih sayang tanpa batas. Tanpa doa dan restu dari mereka, proses penyusunan skripsi ini tidak akan dapat berjalan dengan baik.
2. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT., selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Fetty Tri Anggraeny, S.Kom. M.Kom., selaku Ketua Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Bapak Fawwaz Ali Akbar, S.Kom., M.Kom., selaku dosen pembimbing pertama, yang telah dengan sabar memberikan arahan, saran, dan dukungan selama proses penyusunan skripsi ini.
5. Bapak Achmad Junaidi, S.Kom., M.Kom., selaku dosen pembimbing kedua, atas segala arahan, masukan yang konstruktif, dan bimbingan akademik yang sangat berarti selama proses penelitian ini.
6. Ibu Henni Endah Wahanani, ST, M.Kom., selaku dosen penguji pertama, atas segala kritik dan saran yang membangun selama proses seminar hasil skripsi penelitian ini.
7. Ibu Retno Mumpuni, S.Kom, M.Sc., selaku dosen penguji kedua, atas segala arahan, masukan, serta kritik yang membangun selama proses seminar hasil skripsi penelitian ini.
8. Seluruh dosen Program Studi Informatika, Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur, yang telah memberikan ilmu dan wawasan selama masa perkuliahan.

9. Seluruh sahabat dan rekan seperjuangan, terima kasih atas dukungan, semangat, serta kebersamaan yang tak ternilai selama masa studi dan penyusunan skripsi ini.
10. Pihak-pihak lain yang tidak dapat disebutkan satu per satu, tetapi telah berkontribusi dalam mendukung dan membantu kelancaran penelitian ini.

Penulis menyadari bahwa dalam proses penyusunan skripsi ini terdapat berbagai kekurangan. Oleh karena itu, kritik dan saran yang konstruktif dari semua pihak sangat diharapkan untuk meningkatkan kualitas skripsi ini. Dengan segala keterbatasan yang ada, penulis berharap laporan ini dapat memberikan manfaat baik bagi semua pihak secara umum maupun bagi penulis itu sendiri secara khusus.

Surabaya, 25 Agustus 2025

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	i
LEMBAR PERSETUJUAN	iii
ABSTRAK	vii
KATA PENGANTAR	xi
DAFTAR ISI	xiii
DAFTAR GAMBAR	xv
DAFTAR TABEL	xvii
BAB I PENDAHULUAN	1
1.1 LATAR BELAKANG.....	1
1.2 RUMUSAN MASALAH	4
1.3 TUJUAN PENELITIAN	4
1.4 MANFAAT PENELITIAN	5
1.5 BATASAN MASALAH	5
BAB II TINJAUAN PUSTAKA	7
2.1 PENELITIAN TERDAHULU	7
2.2 Landasan Teori	14
BAB III DESAIN DAN IMPLEMENTASI SISTEM	21
3.1 Metode Penelitian.....	21
3.2 Dataset	22
3.2.1. Data Music Info	22
3.2.2. Data User Listening History	23
3.3 Data Preparation	23
3.3.1. Penanganan Missing Values	24
3.3.2. Deteksi dan Penghapusan Bot.....	25
3.4 Preprocessing Data	26
3.4.1. Transformasi Playcount ke Rating.....	26
3.4.2. Genre Mapping	27
3.5 Clustering	28
3.5.1. Ekstraksi Fitur Genre (TF IDF)	28
3.5.2. Reduksi Dimensi (UMAP) dan T-SNE.....	29
3.5.3 HDBSCAN Clustering	30
3.6 User Item Matrix	31
3.7 Model Matrix Factorization berbasis Alternating Least Squares	33
3.8 Fungsi Rekomendasi	34

3.9	Rekomendasi berbasis Ocasion.....	36
3.10	Evaluasi Model.....	37
3.10.1	RMSE dan MAE.....	37
3.10.2	Reconstruction Error.....	38
3.10.3	Silhouette Score	38
3.11	Skenario Pengujian.....	39
3.11.1	Pengujian Kualitas Clustering	39
3.11.2	Pengujian Kualitas Model Alternating Least Squares	41
3.11.3	Pengujian Kualitas Rekomendasi berdasarkan User Satisfication	42
3.12	Desain Sistem	43
BAB IV PENGUJIAN DAN ANALISA		45
4.1	Metode Pengujian.....	45
4.1.1	Pengujian Pengolahan Data.....	45
4.1.2	Pengujian Kualitas Clustering.....	45
4.1.3	Pengujian Kualitas ALS	45
4.1.4	Pengujian Kepuasan (User Satisfication)	45
4.2	Pengujian Pengolahan Data	46
4.3	Pengujian Perfoma Clustering	50
4.4	Pengujian Performa Model Alternating Least Squares	59
4.4.1	Metode Pengujian Performa Model.....	60
4.4.2	Pembagian Pengujian Performa Model	61
4.4.3	Pengujian Performa dengan Parameter Default	61
4.4.4	Pengujian Performa dengan Variasi Parameter	62
4.4.5	Analisis Hasil Eksperimen	64
4.5	Pengujian Kualitas Rekomendasi berdasar User Satisfication	65
4.5.1	Desain dan Metodologi Pengujian.....	65
4.5.2	Karakteristik Responden	66
4.5.3	Hasil Evaluasi Kualitas Rekomendasi.....	66
4.5.4	Analisis Perkegiatan	67
4.5.5	Tingkat Penyimpanan ke Playlist	67
4.5.6	Feedback Pengguna	68
4.5.7	Implikasi Pengembangan Sistem.....	68
BAB V PENUTUP		71
5.1	Kesimpulan.....	71
5.2	Saran Pengembangan.....	71
DAFTAR PUSTAKA		73

DAFTAR GAMBAR

Gambar 2. 1 Alternating Least Squares [10]	17
Gambar 2. 2 HDBSCAN Clustering [1702.v2]	18
Gambar 2. 3 Integrasi Model Matrix Factorization dan HDBSCAN Clustering	20
Gambar 3. 1 Alur Sistem Rekomendasi	21
Gambar 3. 2 Arsitektur Sistem Rekomendasi	43
Gambar 4. 1 Hasil Mapping Genre	46
Gambar 4. 2 Bobot X Pengujian Ke-1	48
Gambar 4. 3 Bobot X Pengujian Ke-2	49
Gambar 4. 4 Bobot X Pengujian Ke-3	49
Gambar 4. 5 Persebaran dan Nilai Silhouette Score Pengujian Pertama	51
Gambar 4. 6 Visualisasi UMAP Pengujian Pertama	52
Gambar 4. 7 Visualisasi TSNE Pengujian Pertama	52
Gambar 4. 8 Persebaran dan Nilai Silhouette Score Pengujian Kedua	54
Gambar 4. 9 Visualisasi UMAP Pengujian kedua	55
Gambar 4. 10 Visualisasi TSNE Pengujian Kedua	55
Gambar 4. 11 Persebaran dan Nilai Silhouette Score Pengujian Ketiga	57
Gambar 4. 12 Visualisasi Umap Pengujian Ketiga	58
Gambar 4. 13 Visualisasi TSNE Pengujian Ketiga	58
Gambar 4. 14 Pengaruh Parameter n_factor	63
Gambar 4. 15 Pengaruh Parameter regularization	63
Gambar 4. 16 Pertanyaan Kuisisioner	65

DAFTAR TABEL

Tabel 3. 1 Data Music Info	22
Tabel 3. 2 Data User Listening History	23
Tabel 3. 3 Missing Values Filling	24
Tabel 3. 4 Identifikasi dan Penghapusan User Bot.....	25
Tabel 3. 5 Transformasi Playcount ke Rating	27
Tabel 3. 6 Capping Rating.....	27
Tabel 3. 7 Genre Mapping.....	28
Tabel 3. 8 Ekstraksi Fitur Genre.....	29
Tabel 3. 9 Reduksi Dimensi (UMAP)	30
Tabel 3. 10 HDBSCAN Pseudocode	31
Tabel 3. 11 User Per Cluster.....	32
Tabel 3. 12 Normalisasi Matrix.....	32
Tabel 3. 13 Model Per Cluster.....	33
Tabel 3. 14 Fungsi Rekomendasi	34
Tabel 3. 15 Rekomendasi Cold Start User	35
Tabel 3. 16 Rekomendasi (Occasion).....	36
Tabel 3. 17 Reconstruction Error	38
Tabel 3. 18 Silhouette Score.....	39
Tabel 3. 19 Skenario Pengujian Kualitas Clustering.....	40
Tabel 3. 20 Pembagian Data Training dan Validasi.....	41
Tabel 3. 21 Pengujian Performa Model.....	41
Tabel 3. 22 Pengujian Komputasi	41
Tabel 4. 1 User yang terdeteksi sebagai bot.....	47
Tabel 4. 2 Parameter Pengujian Pertama.....	50
Tabel 4. 3 Informasi Cluster Pengujian Pertama	51
Tabel 4. 4 Parameter Pengujian Kedua	53
Tabel 4. 5 Informasi Cluster Pengujian Kedua	53
Tabel 4. 6 Parameter Pengujian Ketiga	56
Tabel 4. 7 Informasi Cluster Pengujian Ketiga	56
Tabel 4. 8 Hasil Evaluasi Model dengan Parameter Default.....	61
Tabel 4. 9 Hasil Eksperimen Variasi Parameter.....	62
Tabel 4. 10 Distribusi Karakteristik Responden.....	66
Tabel 4. 11 Evaluasi Tiga Aspek Utama	66
Tabel 4. 12 Analisis Perkegiatan	67
Tabel 4. 13 Tingkat Penyimpanan berdasarkan Kegiatan	67

