

**Implementasi *Random Forest* untuk Memprediksi Popularitas Lagu di *Spotify*: Studi Kasus  
Fitur Audio dan Metadata**

**Laporan Tugas Akhir**

Diajukan Untuk Memenuhi

Persyaratan Guna Meraih Gelar Sarjana

Informatika Universitas Muhammadiyah Malang



Uswatun Chasanah

(202110370311274)

**Bidang Minat**

Sains Data

**PROGRAM STUDI INFORMATIKA FAKULTAS TEKNIK**

**UNIVERSITAS MUHAMMADIYAH MALANG**

**2024**

**LEMBAR PERSETUJUAN**

**Implementasi *Random Forest* untuk Memprediksi Popularitas Lagu di *Spotify*: Studi Kasus Fitur Audio dan Metadata**

**LEMBAR PERNYATAAN**

**TUGAS AKHIR**

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1  
Prodi Informatika Universitas Muhammadiyah Malang

**Menyetujui**

**Malang, 15 Mei 2025**

**Dosen Pembimbing I**



**Setio Basuki, S.T., M.T., Ph.D.**  
**NIDN-071428403**

## LEMBAR PENGESAHAN

### Implementasi Random Forest untuk Memprediksi Popularitas Lagu di Spotify: Studi Kasus Fitur Audio dan Metadata

### TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1  
Informatika Universitas Muhammadiyah Malang

Disusun Oleh :

**USWATUN CHASANAH**

**202110370311274**

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis pengujian  
pada tanggal 15 Juni 2025

Menyetujui,

Dosen Penguji 1



Christian Sri Kusuma Aditva S.Kom.,

M.Kom

NIP. 180327021991PNS.

Dosen Penguji 2



Ir. Wildan Suharso S.Kom., M.Kom

NIP. 10817030596PNS.

Mengetahui,  
Ketua Jurusan Informatika



Ir. Galih Wasis Wicaksono S.kom. M.Cs.

NIP. 10814100541PNS.

## LEMBAR PERNYATAAN

Yang bertanda tangan di bawah ini :

**NAMA** : Uswatun Chasanah

**NIM** : 202110370311274

**FAK./JUR.** : Fakultas Teknik / Informatika

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul "**Implementasi Random Forest untuk Memprediksi Popularitas Lagu di Spotify Berdasarkan Fitur Audio dan Metadata.**" Beserta seluruh isinya adalah merupakan karya saya sendiri dan bukan merupakan karya tulis orang lain, baik sebagian maupun seluruhnya, kecuali dalam bentuk kutipan yang telah disebutkan berasal dari sumbernya.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini atau ada klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk risiko / sanksi yang berlaku.

Mengetahui,  
Dosen Pembimbing 1

Malang, 15 Mei 2025  
Yang Membuat Pernyataan,



**Setio Basuki, S.T., M.T., Ph.D.**  
**NIDN-071428403**



**Uswatun Chasanah**

## ABSTRAK

Perkembangan industri musik digital yang pesat, khususnya platform *streaming* musik seperti Spotify, mendorong kebutuhan untuk merancang model prediktif untuk memprediksi popularitas lagu. Penelitian ini bertujuan untuk menerapkan algoritma *Random Forest* dalam memprediksi popularitas lagu berdasarkan kombinasi fitur audio dan metadata. Dataset yang digunakan diambil dari Spotify API dengan 2.416 *instance* lagu menggunakan teknik *crawling*, mencakup fitur seperti loudness, tempo, valence, nama artis, dan tanggal rilis. Proses pra-pemrosesan mencakup pembersihan data, encoding, normalisasi, serta seleksi fitur. Model diuji menggunakan berbagai proporsi pembagian data (70:30, 80:20, 60:40, dan 50:50) untuk mengevaluasi performa. Hasil menunjukkan bahwa proporsi 80% dan pelatihan dan 20% pengujian menghasilkan performa terbaik dengan MAE 0.1176, RMSE 0.1526, dan R-square 0.3279. selain itu, hasil analisis interaksi antar fitur yaitu kombinasi seperti *name\_artist\_valence*, *loudness\_acousticness*, dan *length\_acousticness* memberikan kontribusi signifikan terhadap popularitas lagu. Penelitian ini menyimpulkan bahwa kombinasi karakteristik musik dan identitas artis lebih berpengaruh dalam membentuk popularitas lagu dibandingkan waktu rilis.

**Kata Kunci** : Random Forest, Prediksi popularitas Lagu, Fitur Audio, Metadata Lagu, API Spotify, Interaksi Fitur.

## ABSTRACT

The fast-growing digital music industry, especially music streaming platforms like Spotify, has created the need to build models that can predict how popular a song will be. This study aims to use the Random Forest algorithm to predict song popularity based on a mix of audio features and metadata. The dataset was collected using crawling techniques from the Spotify API, with 2,416 songs and features like loudness, tempo, valence, artist name, and release date. The data went through several preprocessing steps, including cleaning, encoding, normalization, and feature selection. The model was tested using different data split ratios (70:30, 80:20, 60:40, and 50:50) to see which gave the best results. The 80% training and 20% testing split showed the best performance, with an MAE of 0.1176, RMSE of 0.01526, and R-squared of 0.3279. The analysis of feature combinations also showed that pairs like name\_artist\_valence, loudness\_acousticness, and length\_acousticness had a strong effect on song popularity. This study concludes that the combination of music features and artist identity has a bigger impact on popularity than the release date.

**Keywords:** Random Forest, Song Popularity Prediction, Audio Features, Song Metadata, Spotify API, Feature Interaction.

## KATA PERSEMBAHAN

Puji syukur saya panjatkan ke hadirat Allah SWT atas segala nikmat dan rahmat-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi ini dengan sebaik-baiknya. Penulis mengucapkan terima kasih kepada:

1. Kedua orang tua saya yang telah berusaha sekuat tenaga, senantiasa mendoakan tanpa henti, menasihati, mendidik, dan menyayangi dengan tulus dan ikhlas.
2. Bapak Setio Basuki, S.T., M.T., Ph.D., selaku dosen pembimbing yang telah meluangkan waktu, tenaga, dan pikiran untuk membimbing penulis dalam menyelesaikan skripsi ini dari awal hingga akhir.
3. Diri saya sendiri, atas keteguhan, semangat belajar, dan perjuangan yang terus dijaga hingga skripsi ini berhasil diselesaikan.
4. Teman-teman semuanya yang telah menemani dan memberikan dukungan selama menempuh perkuliahan.



## KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi ini yang berjudul “**Implementasi *Random Forest* untuk Memprediksi Popularitas Lagu di Spotify Berdasarkan Fitur Audio dan Metadata.**” Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Informatika, Fakultas Teknik, Universitas Muhammadiyah Malang. Adapun topik yang diangkat dalam penelitian ini membahas penerapan algoritma *machine learning*, khususnya *Random Forest*, untuk memprediksi popularitas lagu berdasarkan kombinasi fitur-fitur audio dan metadata platform Spotify. Penelitian ini diharapkan memberikan gambaran tentang bagaimana teknik data science dapat dimanfaatkan dalam industri musik digital untuk mendukung pengambilan keputusan.

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan dan keterbatasan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan untuk perbaikan di masa mendatang. Semoga skripsi ini dapat memberikan manfaat bagi pembaca serta menjadi tambahan wawasan dalam bidang data *science*, khususnya dalam penerapan *machine learning* untuk analisis data musik.

Malang, 1 April 2025



Uswatun Chasanah

# DAFTAR ISI

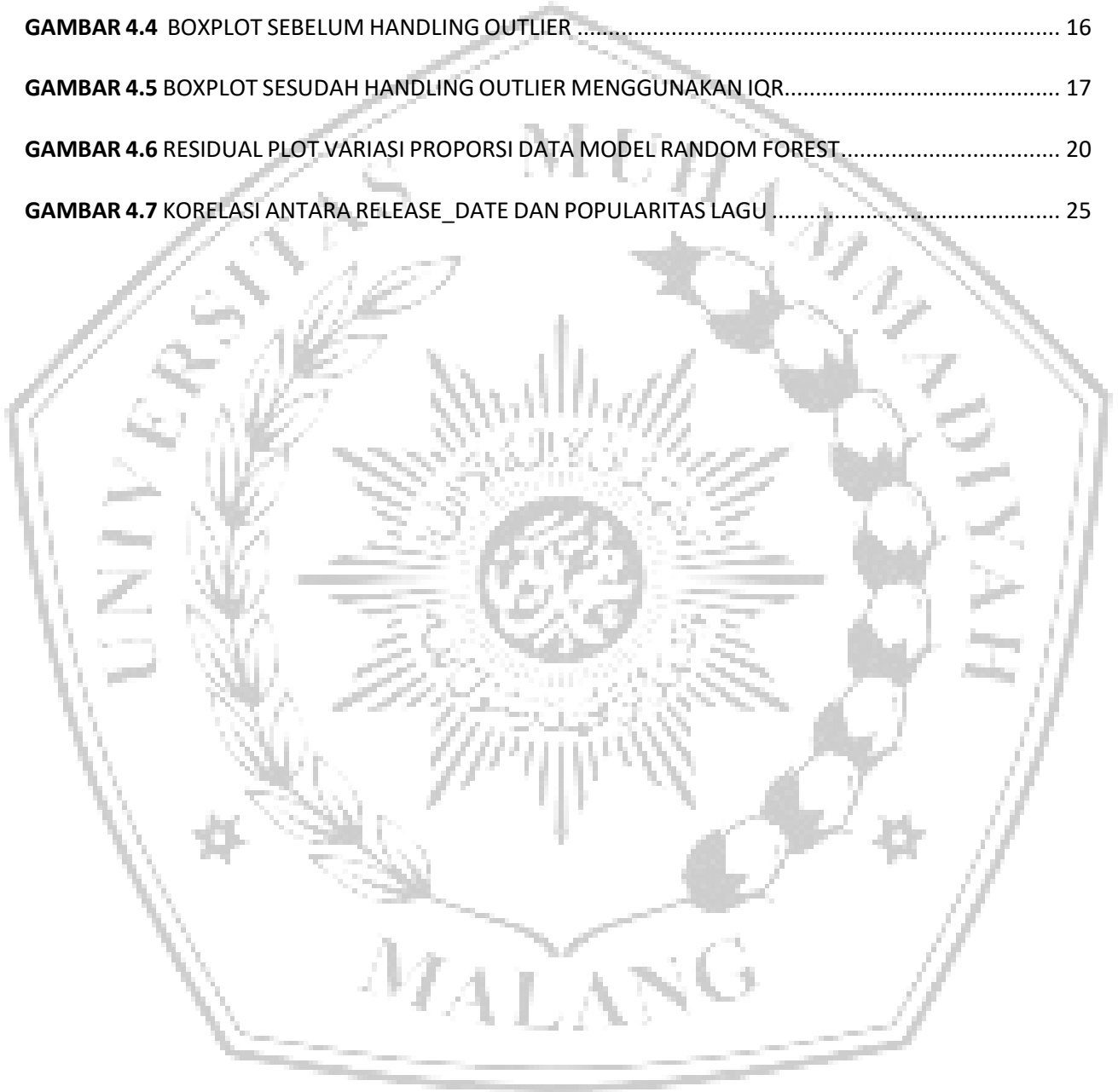
LEMBAR PERNYATAAN .....	ii
DAFTAR TABEL .....	viii
ABSTRAK .....	ix
ABSTRACT .....	x
KATA PENGANTAR.....	xi
KATA PERSEMBAHAN .....	xii
BAB I.....	1
PENDAHULUAN .....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	4
1.3. Tujuan Penelitian .....	4
1.4. Batasan Masalah.....	4
BAB II .....	5
TINJAUAN PUSTAKA .....	5
2.1. Studi Literatur .....	5
2.2. Popularitas Lagu .....	6
2.3. Fitur Audio.....	6
2.4. Fitur Metadata .....	6
2.5. Random Forest Regressor.....	7
BAB III.....	8
METODE PENELITIAN.....	8
3.1. Dataset .....	8
3.2 Pra Pemrosesan Data .....	10
3.3 Pembagian Data.....	11
3.4 Metode Random Forest.....	11
3.5 Evaluasi Model.....	12
3.6 Error Analysis.....	13
BAB IV .....	14
HASIL DAN PEMBAHASAN .....	14
4.1 Hasil Preprocessing Data .....	14
4.2 Penerapan Algoritma random forest dalam prediksi popularitas lagu.....	20
4.3 Analisis interaksi antar fitur terhadap popularitas lagu .....	22
BAB V .....	26
PENUTUP.....	26

5.1 Kesimpulan ..... 26  
5.2 Saran..... 26  
DAFTAR PUSTAKA..... 28



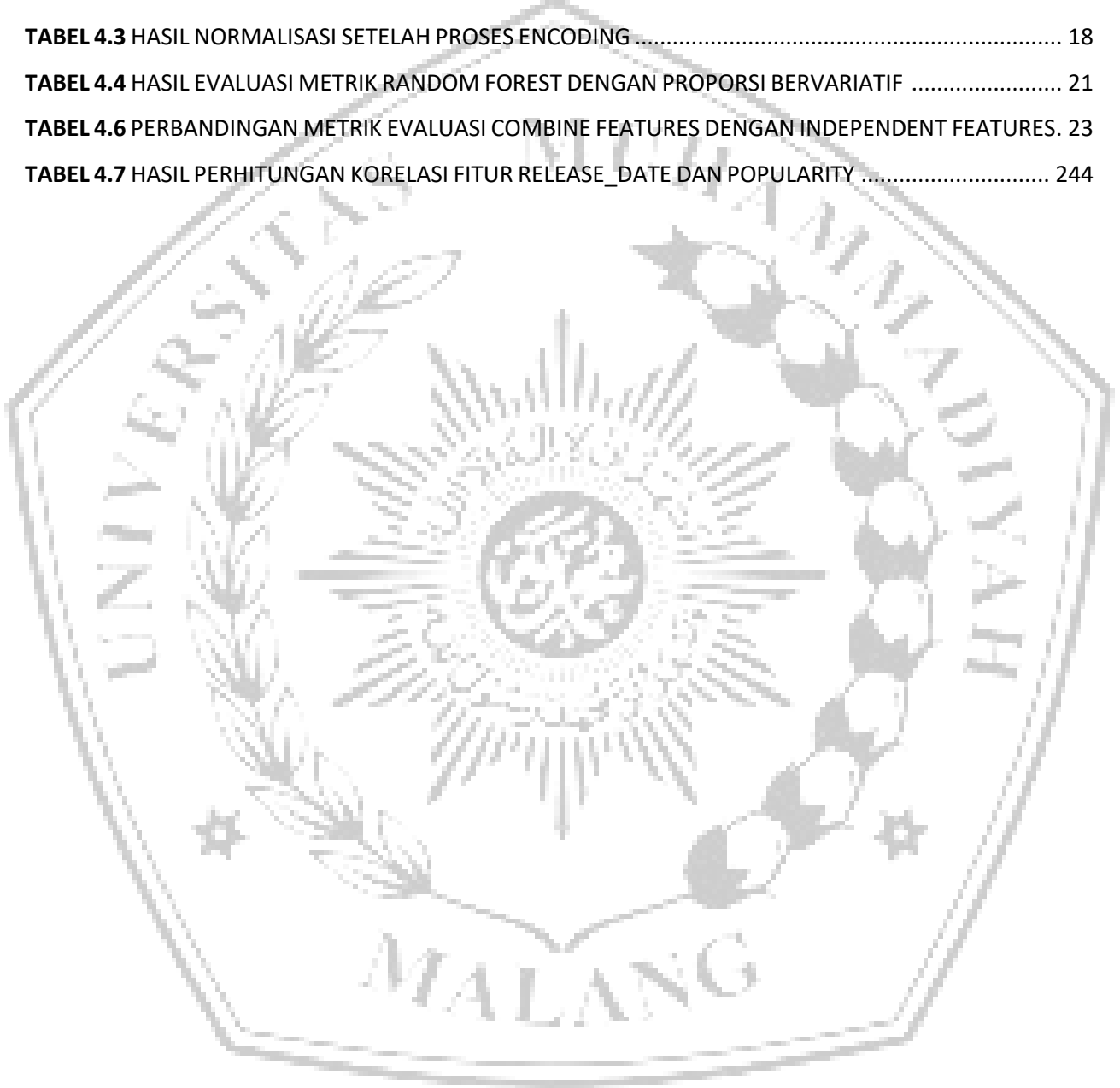
## DAFTAR GAMBAR

<b>GAMBAR 1</b> ALUR PENELITIAN PREDIKSI MENGGUNAKAN RANDOM FOREST .....	8
<b>GAMBAR 4.2</b> HISTOGRAM JUMLAH DUPLIKASI SEBELUM PENGHAPUSAN DUPLIKAT DATA .....	15
<b>GAMBAR 4.3</b> HISTOGRAM JUMLAH DUPLIKASI SETELAH PENGHAPUSAN DUPLIKAT DATA.....	15
<b>GAMBAR 4.4</b> BOXPLOT SEBELUM HANDLING OUTLIER .....	16
<b>GAMBAR 4.5</b> BOXPLOT SESUDAH HANDLING OUTLIER MENGGUNAKAN IQR.....	17
<b>GAMBAR 4.6</b> RESIDUAL PLOT VARIASI PROPORSI DATA MODEL RANDOM FOREST.....	20
<b>GAMBAR 4.7</b> KORELASI ANTARA RELEASE_DATE DAN POPULARITAS LAGU .....	25



## DAFTAR TABEL

<b>TABEL 2.1</b> PENELITIAN TERDAHULU.....	5
<b>TABEL 3.1</b> SAMPEL DATASET DARI API SPOTIFY .....	9
<b>TABEL 4.1</b> SAMPEL DATA TERDETEKSI OUTLIER.....	16
<b>TABEL 4.2</b> HASIL ENCODING KOLOM KATEGORIK .....	18
<b>TABEL 4.3</b> HASIL NORMALISASI SETELAH PROSES ENCODING .....	18
<b>TABEL 4.4</b> HASIL EVALUASI METRIK RANDOM FOREST DENGAN PROPORSI BERVARIATIF .....	21
<b>TABEL 4.6</b> PERBANDINGAN METRIK EVALUASI COMBINE FEATURES DENGAN INDEPENDENT FEATURES. 23	
<b>TABEL 4.7</b> HASIL PERHITUNGAN KORELASI FITUR RELEASE_DATE DAN POPULARITY .....	244



## DAFTAR PUSTAKA

- [1] D. W. Suci, "Manfaat Seni Musik Dalam Perkembangan Belajar Siswa Sekolah Dasar," *Edukatif: Jurnal Ilmu Pendidikan*, vol. 1, no. 3, pp. 177–184, 2019, doi: 10.31004/edukatif.v1i3.45.
- [2] S. Lepa, J. Steffens, M. Herzog, and H. Egermann, "Popular music as entertainment communication: How perceived semantic expression explains liking of previously unknown music," *Media Commun*, vol. 8, no. 3, pp. 191–201, 2020, doi: 10.17645/mac.v8i3.3153.
- [3] Prenika Yuniar, J. K. Sitoena, D. M. Matius, and G. B. Obed, "Sejarah Musik sebagai Dasar Pengetahuan dalam Pembelajaran Teori Musik," *Clef: Jurnal Musik dan Pendidikan Musik*, vol. 3, no. 2, pp. 141–150, 2022, doi: 10.51667/cjmpm.v3i2.1098.
- [4] I. Ruddin, H. Santoso, and R. E. Indrajit, "Digitalisasi Musik Industri: Bagaimana Teknologi Informasi Mempengaruhi Industri Musik di Indonesia Jurnal Pendidikan Sains dan Komputer," *Jurnal Pendidikan Sains dan Komputer*, vol. 2, no. 1, pp. 124–136, 2022.
- [5] U. K-means, S. Marlia, K. Setiawan, and C. Juliane, "Analisis Fitur Musik dan Tren Popularitas Lagu di Spotify menggunakan K - Means dan CRISP - DM Analysis of Music Features and Song Popularity Trends on Spotify," vol. 13, pp. 595–607, 2024.
- [6] L. H. TM, "Peringkat 18 Dunia, Nilai Pasar Musik Streaming di Indonesia Tembus Rp2,1 Triliun." [Online]. Available: <https://teknologi.bisnis.com/read/20200708/84/1262980/peringkat-18-dunia-nilai-pasar-musik-streaming-di-indonesia-tembus-rp21-triliun>
- [7] X. Liu, "Music Trend Prediction Based on Improved LSTM and Random Forest Algorithm," *J Sens*, vol. 2022, 2022, doi: 10.1155/2022/6450469.
- [8] A. I. Journal, "1 , 2 , 3 , 4," vol. 5119, no. 11, pp. 10–19, 2022.
- [9] H. H. Lin, J. Da Lin, J. J. M. Ople, J. C. Chen, and K. L. Hua, "Social Media Popularity Prediction Based on Multi-Modal Self-Attention Mechanisms," *IEEE Access*, vol. 10, pp. 4448–4455, 2022, doi: 10.1109/ACCESS.2021.3136552.
- [10] M. Votter, M. Mayerl, G. Specht, and E. Zangerle, "Novel Datasets for Evaluating Song Popularity Prediction Tasks," *Proceedings - 23rd IEEE International Symposium on Multimedia, ISM 2021*, pp. 166–173, 2021, doi: 10.1109/ISM52913.2021.00034.
- [11] Y. Xu, M. Wang, H. Chen, and F. Hu, "Prediction Model of Music Popular Trend Based on NNS and DM Technology," *Journal of Function Spaces*, vol. 2022, 2022, doi: 10.1155/2022/6104056.
- [12] C. V. S. Araujo, M. A. P. Cristo, and R. Giusti, "A model for predicting music popularity on streaming platforms," *Revista de Informatica Teorica e Aplicada*, vol. 27, no. 4, pp. 108–117, 2020, doi: 10.22456/2175-2745.107021.
- [13] X. Liu, "Music Trend Prediction Based on Improved LSTM and Random Forest Algorithm," vol. 2022, 2022, doi: 10.1155/2022/6450469.
- [14] E. Izquierdo-Verdiguier and R. Zurita-Milla, "An evaluation of Guided Regularized Random Forest for classification and regression tasks in remote sensing," *International*

*Journal of Applied Earth Observation and Geoinformation*, vol. 88, no. June 2019, 2020, doi: 10.1016/j.jag.2020.102051.

- [15] H. S. Saragih, “Predicting song popularity based on Spotify’s audio features: insights from the Indonesian streaming users,” *Journal of Management Analytics*, vol. 10, no. 4, pp. 693–709, doi: 10.1080/23270012.2023.2239824.
- [16] M. Vötter, M. Mayerl, E. Zangerle, and G. Specht, “Song Popularity Prediction using Ordinal Classification,” *Proceedings of the Sound and Music Computing Conferences*, vol. 2023-June, pp. 346–353, 2023.
- [17] A. Çimen and E. Kayış, “A longitudinal model for song popularity prediction,” *Proceedings of the 10th International Conference on Data Science, Technology and Applications, DATA 2021*, no. Data, pp. 96–104, 2021, doi: 10.5220/0010607700960104.
- [18] Z. Xing, “Popularity Prediction of Music by Machine Learning Models,” *Highlights in Science, Engineering and Technology*, vol. 47, pp. 37–45, 2023, doi: 10.54097/hset.v47i.8162.
- [19] E. Zangerle, R. Huber, M. Vötter, and Y. H. Yang, “Hit song prediction: Leveraging low- and high-level audio features,” *Proceedings of the 20th International Society for Music Information Retrieval Conference, ISMIR 2019*, pp. 319–326, 2019.
- [20] A. G. Jondya and B. H. Iswanto, “Analisis dan Seleksi Fitur Audio pada Musik Tradisional Indonesia,” *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi*, vol. 4, no. 2, p. 77, 2018, doi: 10.24014/coreit.v4i2.6506.
- [21] D. Duman, P. Neto, A. Mavrolampados, P. Toiviainen, and G. Luck, “Music we move to: Spotify audio features and reasons for listening,” *PLoS One*, vol. 17, no. 9 September, pp. 1–18, 2022, doi: 10.1371/journal.pone.0275228.
- [22] D. Martin-Gutierrez, G. Hernandez Penaloza, A. Belmonte-Hernandez, and F. Alvarez Garcia, “A Multimodal End-to-End Deep Learning Architecture for Music Popularity Prediction,” *IEEE Access*, vol. 8, pp. 39361–39374, 2020, doi: 10.1109/ACCESS.2020.2976033.
- [23] X. Liu, “Music Trend Prediction Based on Improved LSTM and Random Forest Algorithm,” *J Sens*, vol. 2022, 2022, doi: 10.1155/2022/6450469.
- [24] K. Pawluszek-Filipiak and A. Borkowski, “On the importance of train-test split ratio of datasets in automatic landslide detection by supervised classification,” *Remote Sens (Basel)*, vol. 12, no. 18, 2020, doi: 10.3390/rs12183054.



UNIVERSITAS  
MUHAMMADIYAH  
MALANG



## FAKULTAS TEKNIK

INFORMATIKA  
informatika.umm.ac.id | informatika@umm.ac.id

### FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

Nama Mahasiswa : Uswatun Chasanah

NIM : 202110370311274

Judul TA : Implementasi Random Forest untuk Memprediksi Popularitas Lagu di Spotify: Studi Kasus Fitur Audio dan Metadata

#### Hasil Cek Plagiarisme dengan Turnitin

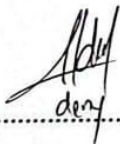
No.	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	9%
2.	Bab 2 – Daftar Pustaka	25 %	20%
3.	Bab 3 – Analisis dan Perancangan	25 %	2%
4.	Bab 4 – Implementasi dan Pengujian	15 %	0%
5.	Bab 5 – Kesimpulan dan Saran	5 %	3%
6.	Makalah Tugas Akhir	20%	0%

\*) Hasil cek plagiarisme diisi oleh pemeriksa (staf TU)

\*) Maksimal 5 kali (4 Kali sebelum ujian, 1 kali sesudah ujian)

Mengetahui,

Pemeriksa (Staff TU)

()



Kampus I  
Jl. Bandung 1 Malang, Jawa Timur  
P. +62 341 551 253 (Hunting)  
F. +62 341 460 435

Kampus II  
Jl. Bendungan Sutarni No 188 Malang, Jawa Timur  
P. +62 341 551 149 (Hunting)  
F. +62 341 582 060

Kampus III  
Jl. Raya Tlogomas No 248 Malang, Jawa Timur  
P. +62 341 464 319 (Hunting)  
F. +62 341 460 435  
E. webmaster@umm.ac.id