

**KLASIFIKASI GENRE MUSIK PADA SPOTIFY DENGAN
MENGUNAKAN XGBOOST**

Skripsi

OLEH:

BIMO ZULFI ANANDA PUTRA

218160031



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MEDAN AREA
MEDAN
2026**

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

Document Accepted 1/7/26

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

KLASIFIKASI GENRE MUSIK PADA SPOTIFY DENGAN MENGUNAKAN XGBOOST

SKRIPSI

**Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana di Fakultas Teknik
Universitas Medan Area**

Oleh:

BIMO ZULFI ANANDA PUTRA

218160031

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MEDAN AREA
MEDAN
2026**

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

Document Accepted 1/7/26

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Access From (repository.uma.ac.id) 1/7/26

HALAMAN PENGESAHAN

Judul Skripsi : Klasifikasi Genre Musik Pada Spotify Dengan Menggunakan XGBoost

Nama : Bimo Zulfi Ananda Putra
NPM : 218160031
Fakultas : Teknik

Disetujui Oleh
Komisi Pembimbing

Andre Hasudungan Lubis, S. Ti., M.Sc
Pembimbing



Pratiomo, S.T, M.T
Kaprodi
Fakultas Teknik



Manono, S.Kom, M.Kom.
Ka. Prodi

Tanggal Lulus: Rabu, 04 Maret 2026

HALAMAN PERNYATAAN

Saya menyatakan bahwa skripsi yang saya susun, sebagai syarat memperoleh gelar sarjana merupakan hasil karya tulis saya sendiri. Adapun bagian-bagian tertentu dalam penulisan skripsi ini yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas sesuai dengan norma, kaidah, dan etika penulisan ilmiah.

Saya bersedia menerima sanksi pencabutan gelar akademik yang saya peroleh dan sanksi-sanksi lainnya dengan peraturan yang berlaku, apabila di kemudian hari ditemukan adanya plagiat dalam skripsi ini.

Medan, 04 Maret 2026



Bimo Zulfi Ananda Putra
218160031

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR/SKRIPSI/TESIS UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Medan Area, saya yang bertanda tangan di bawah ini:

Nama : Bimo Zulfi Ananda Putra
NPM : 218160031
Program Studi : Informatika
Fakultas : Teknik
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Medan Area **Hak Bebas Royalti Noneksklusif (Non-exclusive Royalty-Free Right)** atas karya ilmiah saya yang berjudul :

Klasifikasi Genre Musik Pada Spotify Dengan Menggunakan XGBoost

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Medan Area berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Medan
Pada tanggal : 04 Maret 2026
Yang menyatakan



(Bimo Zulfi Ananda Putra)
218160031

RIWAYAT HIDUP

Bimo Zulfi Ananda Putra lahir di Kota Medan pada tanggal 22 Juni 2003. Penulis merupakan anak ketiga dari tiga bersaudara dari ayah Susanto dan ibunda Maryam. Penulis memulai pendidikan di TK Karya Bunda, Lau Dendang. Kemudian melanjutkan ke jenjang SD Negeri 060853 Sei Kera Hilir II, Kec. Medan Perjuangan sampai kelas 6 dan menyelesaikan pendidikan dasar di sekolah tersebut hingga lulus pada tahun 2015.

Pada tahun yang sama, penulis melanjutkan pendidikan di tingkat menengah pertama, yaitu di SMP Swasta Pertiwi Medan, dan menamatkannya pada tahun 2018. Setelah itu, penulis melanjutkan ke jenjang menengah atas di SMA Negeri 8 Medan sejak tahun 2018 hingga lulus pada tahun 2021. Setelah itu, penulis melanjutkan pendidikan ke perguruan tinggi di Universitas Medan Area pada tahun 2021 hingga tamat di tahun 2026.

KATA PENGANTAR

Puji syukur kita panjatkan kehadiran Allah SWT, yang telah melimpahkan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "**Klasifikasi Genre Musik Pada Spotify Dengan Menggunakan XGBoost**". Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan pendidikan Program Sarjana pada Fakultas Teknik Program Studi Teknik Informatika di Universitas Medan Area. Dalam penyusunan skripsi ini, penulis telah menerima banyak bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada :

1. Allah SWT. yang selalu memberikan kenikmatan dan kesehatan sehingga penulis masih bisa menjalankan hidup dengan baik dan selalu memberikan penulis perlindungan dan kekuatan.
2. Skripsi ini penulis persembahkan untuk kedua orang tua tercinta Ayah dan Ibunda penulis, Susanto dan Maryam yang dengan penuh kasih sayang, ketulusan, dan kerja keras telah berjuang untuk menyekolahkan penulis hingga ke jenjang perkuliahan. Tiap langkah dalam proses penyusunan skripsi ini adalah wujud dari doa dan dukungan kalian.
3. Bapak Prof. Dr. Dadan Ramdan, M.Eng., M.Sc., selaku Rektor Universitas Medan Area.
4. Bapak Dr. Eng. Supriatno, S.T., M.T. selaku Dekan Fakultas Teknik Universitas Medan Area.
5. Bapak Rizki Muliono, S.Kom., M.Kom., selaku Kepala Program Studi Teknik Informatika.

6. Bapak Andre Hasudungan Lubis, S. Ti., M.Sc, selaku dosen pembimbing yang telah meluangkan waktu untuk membimbing dan memberikan arahan kepada penulis dalam proses menyelesaikan skripsi ini.
7. Bapak Dr. Hartono S.Kom., M.Kom., Ibu Nanda Novita S.Kom., M.Kom., dan Ibu Nurul Khairina S.Kom., M.Kom., selaku panitia dalam pelaksanaan pengujian penelitian penulis yang telah memberikan banyak saran dan masukan kepada penulis.
8. Seluruh Dosen Teknik Informatika Universitas Medan Area yang telah memberikan penulis saran yang bermanfaat.
9. Kepada Andini Zahrani, yang selalu memberi motivasi, saya ucapkan Terima kasih atas segala bentuk dukungan, perhatian. Perjalanan menyelesaikan skripsi ini tidak mudah, namun kehadiranmu memberikan kekuatan yang luar biasa. Terima kasih telah selalu ada dalam setiap suka dan duka saya.
10. IT Support Teknik Informatika bang Robby Kurniawan Sari Damanik, ST yang telah membantu menyiapkan pemberkasan yang diperlukan dari awal sampai selesai.
11. Teman-teman kelas penulis yang telah memberikan dukungan dan bantuan selama proses penyusunan skripsi ini.

Medan, 04 Maret 2026



Bimo Zulfi Ananda Putra
218160031

ABSTRAK

Genre musik merupakan salah satu elemen penting dalam mengkategorikan dan merekomendasikan lagu pada platform streaming seperti Spotify. Penelitian ini bertujuan untuk membangun model klasifikasi genre musik menggunakan algoritma XGBoost berdasarkan fitur-fitur numerik dari lagu yang tersedia pada dataset Spotify. Data yang digunakan bersumber dari situs Kaggle, kemudian dipangkas dan difokuskan hanya pada tiga genre utama, yaitu *Pop*, *Rock*, dan *Jazz*. Proses penelitian diawali dengan tahapan praproses data, normalisasi menggunakan *Min-Max Scaling*, dan eksplorasi data melalui statistik deskriptif, histogram, serta analisis korelasi. Model XGBoost diterapkan dengan pengujian parameter *learning rate* (ETA) pada rentang 0,01 hingga 0,1. Hasil terbaik diperoleh pada nilai ETA sebesar 0,05 dengan akurasi mencapai 96%, serta nilai precision, recall, dan f1-score yang tinggi dan seimbang di setiap kelas. Selanjutnya, dilakukan benchmarking terhadap algoritma pembandingan yaitu *K-Nearest Neighbors*, *Naive Bayes*, dan *Decision Tree*. Hasil evaluasi menunjukkan bahwa XGBoost memberikan performa paling unggul dalam seluruh metrik evaluasi. Penelitian ini membuktikan bahwa XGBoost merupakan algoritma yang efektif dan akurat dalam mengklasifikasikan genre musik berdasarkan fitur numerik, serta unggul dibandingkan algoritma klasik lainnya. Temuan ini dapat menjadi dasar dalam pengembangan sistem rekomendasi musik yang lebih cerdas dan personal.

Kata Kunci: Genre Musik, Spotify, XGBoost, Klasifikasi, *Machine Learning*

ABSTRACT

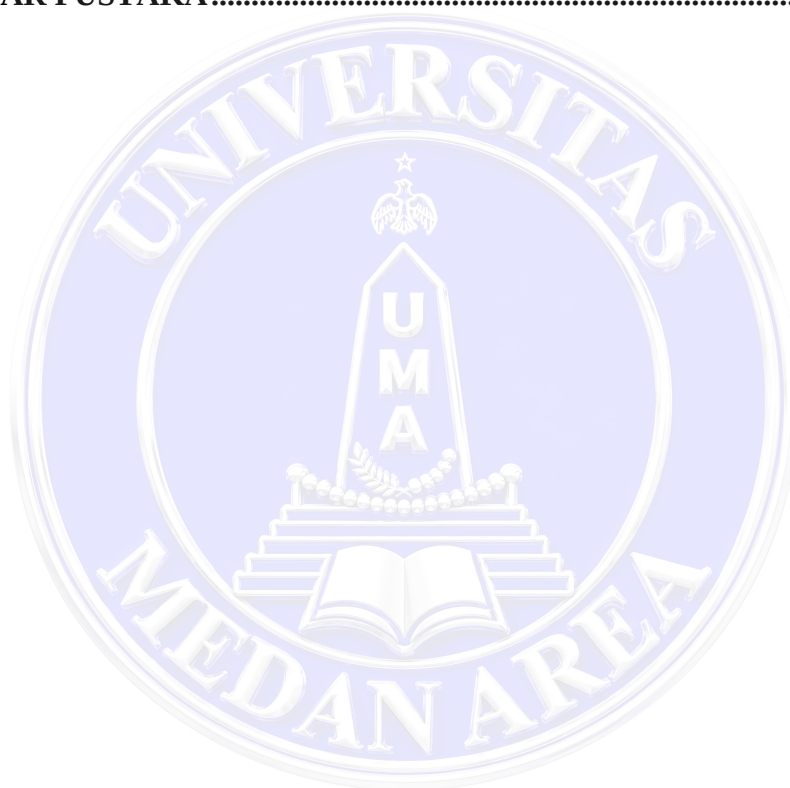
Music genre is a crucial element in categorizing and recommending songs on streaming platforms like Spotify. This study aims to build a music genre classification model using the XGBoost algorithm based on the numerical features of songs available in the Spotify dataset. The data used was sourced from the Kaggle website and then trimmed and focused on only three main genres: Pop, Rock, and Jazz. The research process began with data preprocessing, normalization using Min-Max Scaling, and data exploration through descriptive statistics, histograms, and correlation analysis. The XGBoost model was implemented by testing the learning rate (ETA) parameter in the range of 0.01 to 0.1. The best results were obtained at an ETA of 0.05, achieving 96% accuracy, as well as high and balanced precision, recall, and f1-score values across each class. Next, benchmarking was conducted against comparison algorithms including K-Nearest Neighbors, Naive Bayes, and Decision Tree. The evaluation results showed that XGBoost provided superior performance across all evaluation metrics. This research demonstrates that XGBoost is an effective and accurate algorithm for classifying music genres based on numerical features, outperforming other classical algorithms. These findings can form the basis for developing smarter and more personalized music recommendation systems.

Keyword: Genre Music, Spotify, XGBoost, Classification, Machine Learning

DAFTAR ISI

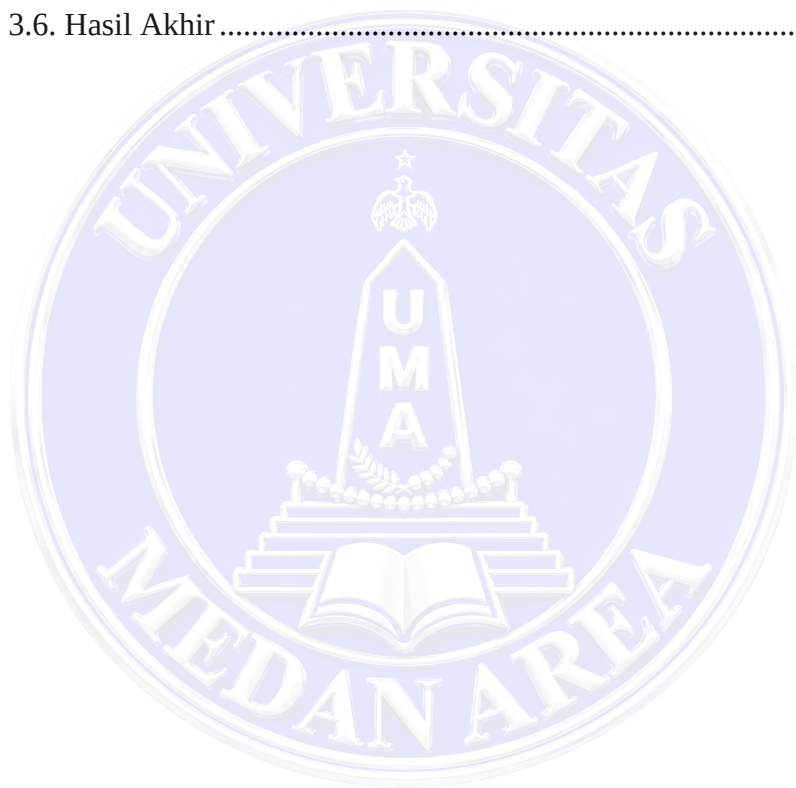
HALAMAN PENGESAHAN.....	i
HALAMAN PERNYATAAN.....	ii
RIWAYAT HIDUP	iv
KATA PENGANTAR.....	v
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
 BAB I PENDAHULUAN.....	 1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	3
1.4. Tujuan Penelitian	4
1.5. Manfaat Penelitian	4
 BAB II TINJAUAN PUSTAKA.....	 6
2.1. Genre Musik.....	6
2.2. Klasifikasi	7
2.3. Boosting	8
2.4. XGBoost.....	9
2.5. Confusion Matrix	11
2.6. Penelitian Terdahulu	12
 BAB III METODOLOGI PENELITIAN	 18
3.1. Tahapan Penelitian	18
3.2. Perhitungan Manual	24
 BAB IV HASIL DAN PEMBAHASAN	 30
4.1. Data yang telah Dikumpulkan	30
4.2. Tahapan Prapemrosesan.....	32
4.2.1. Pemilihan Atribut	32
4.2.2. Pembersihan Data	34
4.2.3. Normalisasi Data	35
4.3. <i>Exploratory Data Analysis</i> (EDA)	37
4.3.1. Statistik Deskriptif	37

4.3.2. Histogram.....	40
4.3.3. Korelasi	43
4.4. Implementasi Extreme Gradient Boosting (XGBoost)	46
4.4.1. <i>Benchmarking</i>	48
4.5. Evaluasi	49
 BAB V KESIMPULAN DAN SARAN	54
5.1. Kesimpulan	54
5.2. Saran.....	54
 DAFTAR PUSTAKA	55



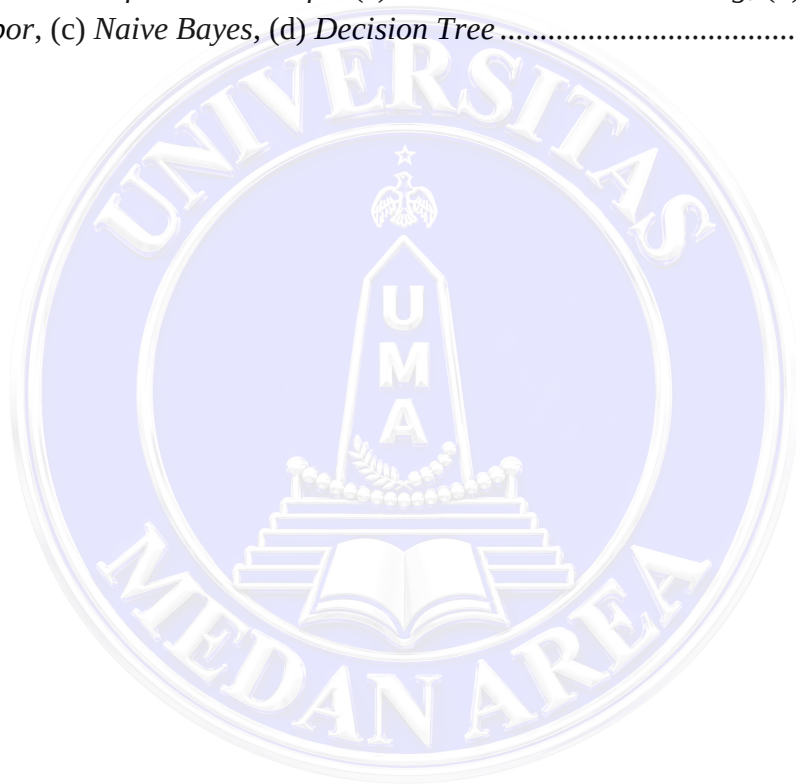
DAFTAR TABEL

Tabel 2.1. Rangkuman Penelitian Terdahulu	12
Tabel 3.1. Rangkuman Dataset	21
Tabel 3.2. Data Dummy	25
Tabel 3.3. Kategori Encoding	25
Tabel 3.4. Setelah Encoding.....	26
Tabel 3.5. <i>Gradient</i> dan <i>Hessian</i>	27
Tabel 3.6. Hasil Akhir	29



DAFTAR GAMBAR

Gambar 3.1 Alur Tahapan Penelitian.....	19
Gambar 4.1 Histogram (a) <i>Popularity</i> , (b) <i>Duration_ms</i> , (c) <i>Key</i> , (d) <i>Loudness</i> , (e) <i>Acousticness</i> , (f) <i>Instrumentalness</i> , (g) <i>Tempo</i> , (h) <i>Time_Signature</i> , (i) <i>Danceability</i> , (j) <i>Energy</i> , (k) <i>Mode</i> , (l) <i>Speechiness</i> , (m) <i>Liveness</i> , (n) <i>Valence</i> , (o) <i>Track_Genre</i> pada Data Sekunder	42
Gambar 4.2 <i>Heatmap</i> Korelasi Data Sekunder	44
Gambar 4.3 <i>Confusion Matrix</i> for (a) <i>Extreme Gradient Boosting</i> , (b) <i>K-Nearest Neighbor</i> , (c) <i>Naive Bayes</i> , (d) <i>Decision Tree</i>	52



BAB I

PENDAHULUAN

1.1. Latar Belakang

Musik telah menjadi cara untuk mempengaruhi orang sejak dahulu. Musik dapat memberikan dampak yang berbeda, mulai dari menenangkan hingga membangkitkan semangat. Setiap orang merespons musik dengan cara yang unik, tergantung pada kondisi emosional dan pengalaman hidup mereka. (Lindholm, 2022). Spotify merupakan sebuah platform *streaming* audio yang memiliki ratusan juta pengguna dengan mencakup jutaan lagu dan *podcast* sebagai layanan kepada pengguna (De Nadai dkk., 2024). Spotify memungkinkan pengguna untuk mendengarkan musik secara fleksibel dan kapan saja di mana saja melalui perangkat elektronik seperti ponsel, komputer, atau tablet. Sehingga, Spotify memfasilitasi pengalaman mendengarkan musik secara praktis dan personal, serta mendukung hubungan antara musik, identitas pengguna, dan pengalaman pengguna dalam kehidupan sehari-hari (Demir, 2022). Penggunaan platform streaming musik semakin meningkat, dan hal ini secara besar mengubah selera, kebiasaan, serta sikap para pendengar (Barate dkk., 2022).

Saat ini telah banyak alat atau fitur yang memudahkan pengguna dalam mendengarkan musik, salah satunya adalah sistem rekomendasi. Proses rekomendasi musik melibatkan berbagai kriteria, antara lain nama artis atau penyanyi, jenis album atau *single*, tahun rilis, label genre musik, serta *playlist* yang

dipilih secara khusus. Hal ini memungkinkan pengalaman mendengarkan musik yang lebih personal dan mudah sesuai preferensi masing-masing pengguna (Demir, 2022). Ukuran besar katalog musik mereka telah mendorong pengembangan berbagai alat inovatif yang membantu pengguna menemukan lagu yang paling sesuai dengan selera mereka di antara begitu banyak pilihan. Kebanyakan alat ini menganalisis profil dan kebiasaan mendengarkan pengguna dengan menggunakan teknik kecerdasan buatan, seperti *collaborative filtering* atau *content-based filtering*, lalu memberikan rekomendasi musik yang bersifat personal kepada pengguna (Álvarez dkk., 2023).

Extreme Gradient Boosting (XGBoost), merupakan sebuah teknik penggabungan *decision tree-based*. XGBoost bekerja dengan meminimalkan kesalahan model dan meningkatkan kemampuan prediksi dengan menggabungkan beberapa model lemah menjadi satu model yang lebih kuat (Noh dkk., 2021). Algoritma XGBoost diperkenalkan oleh Chen dan Guestrin, dan menjadi salah satu pendekatan yang sangat efektif dalam bidang *Machine Learning*. Banyak penelitian yang berhasil menggunakan model XGBoost untuk memecahkan masalah prediksi deret waktu, seperti memprediksi harga saham, energi, demam berdarah, harga minyak, dan lalu lintas kendaraan. Tidak ketinggalan, masalah perkiraan beban listrik juga banyak diteliti oleh para ilmuwan menggunakan model XGBoost, dan hasilnya cukup mengesankan (Tran dkk., 2023). Hal ini karena XGBoost memiliki keunggulan seperti hasil keputusan yang baik, kecepatan komputasi yang cepat, portabilitas yang kuat, dan mampu terintegrasi dengan *cloud*. Selain itu, fitur pentingnya yang dihasilkan oleh XGBoost dapat membantu menjelaskan hubungan

antara hasil *output* dan fitur *input*, sehingga operator jaringan bisa lebih memahami hasil deteksi kegagalan atau kerusakan (C. Zhang dkk., 2021).

Algoritma XGBoost memiliki keunggulan dibandingkan algoritma tradisional dalam mengolah data yang tidak seimbang (*imbalanced data*) (Y. Zhang & Chen, 2021). kemampuannya untuk membuat pohon keputusan yang memiliki banyak *output* sekaligus, berbeda dengan model pohon tradisional yang biasanya hanya memiliki satu output (Kiriakidou dkk., 2024). Dibandingkan dengan algoritma *Machine Learning* lain seperti *Random Forest* dan *Support Vector Machines*, XGBoost lebih banyak digunakan karena memiliki kemampuan dalam hal kecepatan dan akurasi (Qin, 2022). Sehingga, penelitian ini akan melakukan klasifikasi genre musik pada data dari Spotify dengan menggunakan algoritma XGBoost.

1.2. Rumusan Masalah

Penelitian ini merumuskan masalah sebagai berikut: Bagaimana mengklasifikasi genre musik pada platform Spotify dengan menggunakan XGBoost.

1.3. Batasan Masalah

1. Penelitian ini berfokus pada klasifikasi Genre Musik yang di sediakan oleh platform Spotify.
2. Model yang akan dibangun pada penelitian ini adalah menggunakan Algoritma *Extreme Gradient Boosting* (XGBoost).

3. Dataset yang digunakan bersifat data sekunder yang berasal dari Kaggle dengan tautan yakni: ([Spotify Tracks Genre](#)).
4. Adapun atribut pada data penelitian berjumlah sebanyak 15, diantaranya artist, judul lagu, album, popularitas, dan eksplisit.
5. Evaluasi kinerja model menggunakan *Confusion Matriks* dengan parameter akurasi, presisi, *recall*, dan *F-1 score*.

1.4. Tujuan Penelitian

Membangun model klasifikasi dengan menggunakan Algoritma XGBoost untuk melakukan klasifikasi genre musik pada preferensi selera pengguna.

1.5. Manfaat Penelitian

Berikut adalah manfaat penelitian dari perspektif Praktis, Metodologis, dan Teoritis:

1. Praktis:

Dengan klasifikasi genre musik yang lebih akurat, sistem rekomendasi di platform seperti Spotify dapat menyarankan lagu yang lebih sesuai dengan preferensi pengguna, meningkatkan kepuasan dan keterlibatan pengguna.

2. Metodologis:

Penelitian ini memungkinkan perbandingan kinerja XGBoost dengan algoritma lain seperti SVM dan Random Forest dalam klasifikasi genre musik, memberikan wawasan tentang kelebihan dan kekurangan masing-masing model.

3. Teoritis:

Hasil penelitian dapat digunakan untuk mengembangkan model klasifikasi genre musik yang lebih akurat dan efisien, serta memberikan dasar untuk penelitian lebih lanjut dalam bidang ini.



BAB II

TINJAUAN PUSTAKA

2.1. Genre Musik

Musik adalah bidang di mana proses belajar diharapkan terjadi melalui persepsi dan diskriminasi. Selama proses ini, bidang musik memiliki struktur yang kompleks, yang membuatnya sulit untuk dipahami dan memahami konsep yang terdiri dari banyak elemen ini. Salah satu metode yang paling efektif untuk mempertahankan keteraturan ini adalah dengan mengkategorikan genre musik (Ceylan dkk., 2021). Aspek terpenting dalam pengambilan informasi musik adalah pengenalan genre musik. Dalam beberapa tahun terakhir, pengenalan genre musik menjadi bidang penelitian yang sangat penting (Kostrzewa dkk., 2021). Selain itu, persepsi manusia terhadap musik, yang dipengaruhi oleh opini dan pengalaman pribadi, membuat kesepakatan tentang genre menjadi semakin sulit (Castillo & Flores, 2021).

Langkah pertama untuk semua tugas seperti ini adalah mengumpulkan data tentang karakteristik lagu dan metadata yang relevan tentang dunia musik (misalnya, popularitas, data tentang artis dan genre) (Oliveira dkk., 2022). Meskipun genre tersebut tidak disebutkan secara eksplisit. Di halaman utama, Spotify menampilkan interaksi terakhir pengguna dan playlist yang dibuat pengguna sendiri (Hesmondhalgh dkk., 2024). Meskipun kategori analisis musik sangat efektif untuk memahami musik dalam beberapa gaya musik, penting untuk

menyadari bahwa kategori tersebut tidak selalu relevan, atau mungkin tidak cocok secara langsung dengan semua jenis musik (Deruty dkk., 2022).

2.2. Klasifikasi

Klasifikasi adalah proses dalam *machine learning* yang bertujuan untuk mengelompokkan data ke dalam salah satu dari beberapa kelas atau kategori yang telah ditentukan (Chaudhury dkk., 2022). Contohnya adalah mengklasifikasikan *email* ke dalam kategori “spam” atau “bukan spam”, atau dalam konteks penelitian ini, mengklasifikasikan lagu-lagu di Spotify ke dalam genre tertentu berdasarkan fitur-fitur musik yang tersedia (Sari & Mahalisa, 2024). Model klasifikasi belajar dari data pelatihan yang telah diberi label, lalu mengaplikasikan pola yang ditemukan untuk mengklasifikasikan data baru (Strelcenia & Prakoonwit, 2023).

Cara kerja klasifikasi dimulai dari proses pelatihan (*training*), di mana model mempelajari hubungan antara fitur-fitur input (seperti tempo, energi, *valence*, dll.) dan label kelas (seperti genre musik). Setelah pelatihan selesai, model diuji dengan data yang belum pernah dilihat sebelumnya (*testing*) untuk mengukur kemampuannya dalam melakukan prediksi (Kufel dkk., 2023). Berbagai algoritma dapat digunakan dalam klasifikasi, seperti *decision tree*, *support vector machine* (SVM), *k-nearest neighbor* (KNN), dan XGBoost (Naufal, 2021). Algoritma-algoritma ini memiliki keunggulan dan kelemahannya masing-masing tergantung pada karakteristik data yang digunakan (Moshawrab dkk., 2023).

Machine learning atau pembelajaran mesin adalah cabang dari kecerdasan buatan (*artificial intelligence*) yang berfokus pada pengembangan algoritma yang memungkinkan komputer belajar dari data tanpa harus diprogram secara eksplisit

(Maesaroh dkk., 2024). Sistem *machine learning* bekerja dengan menganalisis data dan mengenali pola, sehingga dapat membuat keputusan atau prediksi berdasarkan data baru yang serupa (Pratiwi dkk., 2025). Teknologi ini banyak diterapkan dalam berbagai bidang, seperti pengenalan wajah, deteksi penipuan, rekomendasi konten, dan klasifikasi data (EL Fadel, 2025).

Machine learning secara umum dibagi menjadi dua kategori utama, yaitu *supervised learning* dan *unsupervised learning* (Nurhalizah dkk., 2024). *Supervised learning* bekerja dengan data yang telah diberi label, di mana model belajar dari pasangan *input-output* untuk memprediksi label pada data baru. Sebaliknya, *unsupervised learning* digunakan untuk menemukan pola atau struktur dalam data yang belum diberi label, seperti dalam teknik *clustering* atau asosiasi (Indonesia, 2025). Dalam konteks ini, klasifikasi merupakan bagian dari *supervised learning*, karena model dilatih menggunakan data yang telah memiliki label kelas tertentu (Sabiri dkk., 2024).

2.3. Boosting

Boosting adalah salah satu teknik pembelajaran ansambel (*ensemble learning*) dalam *machine learning* yang bertujuan meningkatkan performa model dengan menggabungkan beberapa *weak learner* menjadi satu *strong learner*. *Weak learner* adalah model prediksi sederhana yang kinerjanya hanya sedikit lebih baik dari tebakan acak, seperti *decision tree* dangkal (Cendani & Wibowo, 2022). Dalam *boosting*, model-model tersebut dilatih secara berurutan, di mana setiap model baru difokuskan untuk memperbaiki kesalahan prediksi yang dilakukan oleh model

sebelumnya (Turaina & Saputra, 2024). Dengan cara ini, boosting secara bertahap meningkatkan akurasi model secara keseluruhan (Ginting dkk., 2022).

Salah satu algoritma boosting yang paling populer dan banyak digunakan adalah XGBoost (*Extreme Gradient Boosting*). XGBoost merupakan implementasi lanjutan dari teknik *gradient boosting* yang dirancang agar efisien dan akurat, dengan fitur tambahan seperti regularisasi, pemrosesan paralel, dan pruning pohon keputusan (Ismanto & Novalia, 2021). Dalam konteks klasifikasi genre musik di Spotify, XGBoost memanfaatkan fitur-fitur audio seperti tempo, energi, dan valensi untuk membangun model prediksi yang kuat (Tannady dkk., 2024). Dengan menekankan pada data yang sulit diklasifikasikan selama proses pelatihan, boosting memungkinkan model mengenali pola-pola yang kompleks dalam data musik dan menghasilkan prediksi genre yang lebih akurat dan andal (Wairata dkk., 2021).

2.4. XGBoost

XGBoost (*Extreme Gradient Boosting*) adalah algoritma pembelajaran ansambel berbasis pohon keputusan yang dirancang untuk memberikan kinerja tinggi dan efisiensi dalam pemrosesan data (Effectiveness & Regression, 2024). Algoritma ini merupakan pengembangan dari metode *gradient boosting*, dengan berbagai optimasi seperti regularisasi, pengolahan paralel, dan kontrol *overfitting* yang lebih baik (Suryana dkk., 2021). Karena kemampuannya dalam mengelola data dengan fitur kompleks dan volume besar, XGBoost menjadi salah satu algoritma yang populer dalam berbagai kompetisi dan aplikasi machine learning, termasuk klasifikasi genre musik (Randi Farmana Putra dkk., 2021).

Secara umum, XGBoost bekerja dengan membangun model prediksi secara bertahap, di mana setiap model baru difokuskan untuk memperbaiki kesalahan prediksi dari model sebelumnya. Proses ini dilakukan dengan meminimalkan fungsi objektif yang terdiri dari dua bagian utama, yaitu fungsi *loss* dan fungsi regularisasi (Septiana Rizky dkk., 2022). Fungsi objektif tersebut dituliskan sebagai berikut:

$$\mathcal{L}(\phi) = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \sum_{k=1}^t \Omega(f_k) \quad (2.1)$$

Pada rumus di atas, $l(y_i, \hat{y}_i^{(t)})$ adalah fungsi *loss* yang mengukur selisih antara label asli dan prediksi, sedangkan $\Omega(f_k)$ merupakan fungsi regularisasi yang digunakan untuk mengontrol kompleksitas model. Regularisasi ini penting untuk menghindari *overfitting*, yaitu kondisi di mana model terlalu cocok dengan data pelatihan tetapi buruk dalam memprediksi data baru (Tantyoko dkk., 2023). Fungsi regularisasi dalam XGBoost biasanya dirumuskan sebagai:

$$\Omega(f) = \gamma T + \frac{1}{2} \sum_{j=1}^T w_j^2 \quad (2.2)$$

Dalam konteks klasifikasi genre musik di Spotify, XGBoost digunakan untuk memetakan fitur-fitur audio seperti tempo, *energy*, *valence*, dan ke dalam label genre yang sesuai (LUTFIANI dkk., 2023). Kemampuan XGBoost dalam menangani hubungan *non-linear* antar fitur menjadikannya sangat efektif dalam menemukan pola yang tersembunyi dalam data musik (Setiawan & Rahmatulloh, 2025). Dengan kombinasi antara akurasi tinggi dan kecepatan pelatihan, XGBoost merupakan pilihan yang tepat untuk menyelesaikan tugas klasifikasi dengan banyak variabel input seperti pada data musik digital (Tuning, 2025).

2.5. Confusion Matrix

Confusion matrix merupakan salah satu metode evaluasi yang paling mendasar dan penting dalam menilai performa model klasifikasi dalam *machine learning* (Fadli & Saputra, 2023). Matriks ini menyajikan informasi tentang bagaimana model membagi prediksi ke dalam kategori yang benar dan salah, baik untuk kelas positif maupun negatif (Mousavi dkk., 2024). Dalam bentuk tabel dua dimensi, *confusion matrix* memungkinkan pengguna untuk mengamati secara langsung sejauh mana model berhasil mengklasifikasikan data dengan benar dan di mana model melakukan kesalahan. Ini sangat bermanfaat terutama ketika data terdiri dari lebih dari dua kelas atau ketika terjadi ketidakseimbangan jumlah data antar kelas (Barranco-Chamorro & Carrillo-García, 2021).

Salah satu metrik evaluasi yang paling umum dan mudah dipahami adalah akurasi (*accuracy*). Akurasi mengukur seberapa besar proporsi prediksi model yang benar dibandingkan dengan total seluruh data yang diuji (Epstein dkk., 2023). Secara matematis, rumus akurasi dapat dituliskan sebagai berikut (Simanjuntak dkk., 2023):

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN} \quad (2.3)$$

$$Precision = \frac{TP}{TP + FP} \quad (2.4)$$

$$Recall = \frac{TP}{TP + FN} \quad (2.5)$$

$$F1-Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (2.6)$$

Akurasi memberikan gambaran umum mengenai kinerja model. Jika akurasi tinggi, maka model secara keseluruhan mampu membuat prediksi yang benar dalam sebagian besar kasus (Safitri & Arifin Zaenal, 2023). Namun, perlu diperhatikan bahwa akurasi bisa menjadi metrik yang menyesatkan apabila data bersifat tidak seimbang (misalnya, satu kelas jauh lebih dominan dibandingkan kelas lainnya) (Nurussakinah & Faisal, 2023). Dalam kasus seperti ini, model dapat menghasilkan akurasi tinggi hanya dengan memprediksi kelas mayoritas, meskipun sebenarnya gagal mengenali kelas minoritas (Meléndez dkk., 2024).

2.6. Penelitian Terdahulu

Terdapat berbagai penelitian yang menggunakan berbagai metode dan *object* yang berbeda. Berikut rangkuman dari penelitian terdahulu yang dapat dilihat pada tabel dibawah ini:

Tabel 2.1. Rangkuman Penelitian Terdahulu

Penulis (tahun)	Judul	Objek	Metode	Hasil
(Saputro & Alamsyah, 2024)	Perbandingan Naive Bayes Classifier dan Algoritma K-Nearest Neighbor dengan Information Gain dan Adaptive Boosting untuk Analisis Sentimen Ulasan Aplikasi Spotify	Ulasan Aplikasi Spotify	Preprocessing Data, Lexicon based labelling with TextBlob, Normalization, Word Vectorization menggunakan TF-IDF, Naive Bayes Classifier, K-Nearest Neighbor	Classifier Naïve Bayes yang menggunakan teknik information gain dan adaptive boosting mencapai 87,28%. Sementara itu, metode k-nearest neighbor yang juga menggunakan information gain dan adaptive boosting memperoleh akurasi sebesar 80,35%. Hasil ini didapatkan dengan menggunakan

				dataset sebanyak 60.000 data, yang dibagi menjadi 80% data untuk pelatihan dan 20% data untuk pengujian.
(Musyarofah dkk., 2022)	Klasifikasi Top 50 Spotify Tahun 2010-2019 Menggunakan Metode K-Means Clustering	Top 50 Spotify Tahun 2010-2019	K-Means Clustering	Hasil kluster kriteria tertinggi genre Pop ditentukan berdasarkan valence lagu. Valence lagu pada Spotify berarti semakin tinggi nilai valensi atau titik akuratnya, maka lagu tersebut memiliki pesan bahagia (nada mayor). Valensi lagu dapat mempengaruhi pendengar, pendengar menjadi lebih senang dan bahagia. Maka dari itu, kluster diagram batang Genre Pop dari Top Popular artis menunjukkan memiliki tingkat valensi yang tinggi. Posisi kedua diperoleh kriteria Beats Per Minute (BPM), bahwa Genre Pop lagu Popular artis dipengaruhi oleh ketukan/tempo lagu untuk memastikan cepat atau lambatnya lagu tersebut.

(Tahiyat dkk., 2025)	Klasterisasi Lagu Populer dan Eksplorasi Subgenre Spotify 2024 dengan K-Medoids	Lagu Populer dan Eksplorasi Subgenre Spotify 2024	K-Medoids	Hasil pengelompokan menggunakan algoritma K-Medoids menunjukkan bahwa lagu lagu dalam dataset dapat dikelompokkan menjadi tiga klaster berdasarkan Spotify Playlist Count, Spotify Playlist Reach, dan Spotify Popularity. Klasterisasi dilakukan dalam dua tahap, yaitu tanpa pengurutan popularitas dan setelah dilakukan pengurutan popularitas, guna me mahami bagaimana distribusi popularitas lagu memengaruhi pola klasterisasi.
(Hendri dkk., 2024)	Perbandingan Algoritma Decision Tree, Naive Bayes, K-Nearest Neighbors pada Genre Musik Spotify	Genre Musik Spotify	Decision Tree, Naive Bayes, K-Nearest Neighbors	Hasil penelitian menunjukkan bahwa algoritma Decision Tree menghasilkan tingkat akurasi sebesar 23%, Naive Bayes 17%, dan K-Nearest Neighbors 19%. Penelitian ini memberikan gambaran tentang preferensi pendengar musik di Spotify dalam memilih genre musik. Berdasarkan hasil penelitian, algoritma Decision Tree memiliki tingkat akurasi tertinggi dalam

				mengklasifikasikan genre musik di Spotify, dengan genre musik dance elektronik (EDM) menjadi yang paling populer di kalangan pengguna Spotify, diikuti oleh rap, pop, R&B, Latin, dan rock. Sementara itu, algoritma Naive Bayes dan K-Nearest Neighbors menunjukkan tingkat keakuratan yang lebih rendah. Penelitian ini memberikan wawasan penting tentang pilihan genre musik bagi pengguna Spotify, dengan Decision Tree sebagai pilihan yang paling akurat. Hasil ini dapat membantu pendengar musik dalam mendapatkan rekomendasi musik yang sesuai di platform Spotify.
(Navisa dkk., 2021)	Komparasi Algoritma Klasifikasi Genre Musik pada Spotify Menggunakan CRISP-DM	Genre Musik Spotify	CRISP-DM	Hasil dari penelitian dan pembahasan terkait komparasi klasifikasi genre musik menggunakan Cross Industry Standard Process for Data Mining (CRISP-DM) dapat disimpulkan bahwa Klasifikasi genre musik pada Spotify menggunakan algoritma klasifikasi yaitu

				Naive Bayes, K-NN dan Random Forest. Validasi menggunakan 10 fold cross validation dan uji menggunakan t-test. Performa dari ketiga algoritma kemudian dibandingkan, mana algoritma yang memiliki hasil performa yang baik. Pengujian dilakukan dengan dataset of song in Spotify, yang terdiri dari 42.305 data. Hasil tertinggi nilai akurasi didapatkan oleh algoritma klasifikasi Naive Bayes yaitu sebesar 58.91%.
	Breast Cancer Classification using XGBoost	Breast Cancer	Extreme Gradient Boosting (XGBoost)	Kami telah menggunakan algoritma pembelajaran mesin Extreme Gradient Boosting (XGBoost) dalam makalah ini. Dengan menerapkan algoritma ini, kami memperoleh akurasi sebesar 94,74% untuk XGBoost. Kami akan mencoba memperkuat penelitian kami di masa depan dengan menangani dataset yang relatif lebih besar dan menggabungkan

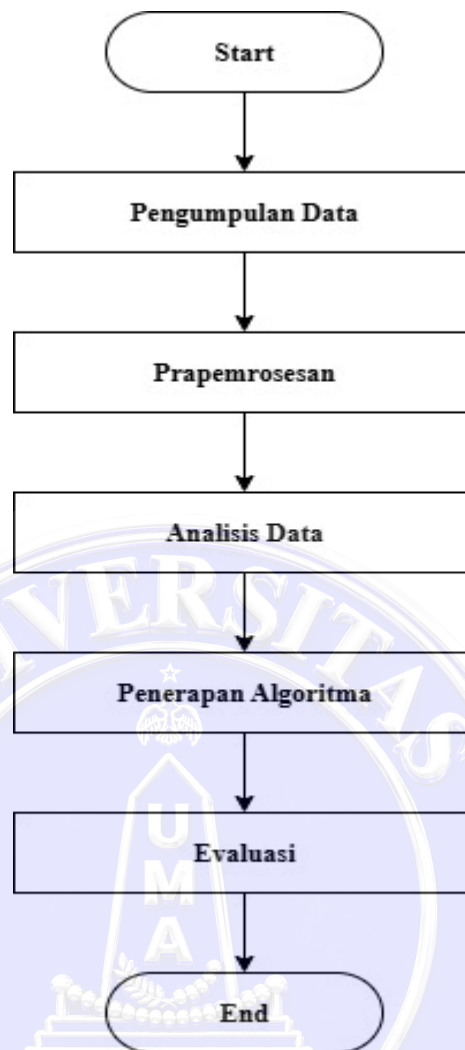
				beberapa fungsi tambahan seperti deteksi fase kanker payudara dan sebagainya. Kami berharap studi ini akan berkontribusi pada aplikasi klinis dalam pengobatan kanker payudara.
	Deep Learning and Machine Learning with Grid Search to Predict Later Occurrence of Breast Cancer Metastasis Using Clinical Data	Breast Cancer	Deep FeedForward Neural Network, Naïve Bayes, Logistic Regression, Support Vector Machine, LASSO, Decision Tree, K-Nearest Neighbor, Random Forest, AdaBoost, dan Extreme Gradient Boosting.	Hasil kami menunjukkan bahwa pembelajaran mendalam dengan grid search secara keseluruhan berkinerja setidaknya sama baiknya dengan metode pembelajaran mesin lainnya ketika menggunakan data klinis non-citra. Menarik untuk dicatat bahwa beberapa metode pembelajaran mesin lainnya, seperti XGB, RF, dan SVM, merupakan pesaing yang sangat kuat bagi DFNN ketika menggabungkan grid search. Perlu juga dicatat bahwa waktu komputasi yang dibutuhkan untuk melakukan grid search dengan DFNN jauh lebih lama daripada yang dibutuhkan untuk melakukan grid search dengan sembilan metode pembelajaran mesin lainnya.

BAB III

METODOLOGI PENELITIAN

3.1. Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahapan yang bertujuan untuk menganalisis performa algoritma XGBoost dalam melakukan klasifikasi genre musik berdasarkan fitur-fitur audio dari Spotify. Tahapan pertama dimulai dengan proses pengumpulan data, yaitu dengan mengakses dataset Spotify yang berisi informasi musik seperti tempo, *valence*, dan fitur audio lainnya dari situs Kaggle. Setelah data diperoleh, dilakukan tahap *preprocessing* untuk membersihkan dan mempersiapkan data agar siap digunakan dalam pelatihan model. Selanjutnya, dilakukan pelatihan model menggunakan algoritma XGBoost, kemudian dievaluasi menggunakan metrik evaluasi seperti akurasi untuk mengukur kinerja model dalam mengklasifikasikan genre musik. Berikut adalah *flowchart* yang menggambarkan alur tahapan dalam penelitian ini:



Gambar 3.1 Alur Tahapan Penelitian

3.1.1. Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan dengan memanfaatkan dataset sekunder yang tersedia secara publik melalui platform Kaggle. Dataset yang Penelitian menggunakan berjudul "Spotify Tracks Genre" yang dapat diakses di “<https://www.kaggle.com/datasets/thedevastator/spotify-tracks-genre-dataset>”.

Pemilihan dataset ini dikarenakan memenuhi beberapa kriteria penting yang mendukung tujuan dan metodologi penelitian ini. Dataset ini mencakup lebih dari 114.000 lagu dari berbagai genre, yang menyediakan representasi luas dari musik yang tersedia di platform Spotify. Data ini mencakup berbagai fitur audio seperti

danceability, energy, loudness, tempo, acousticness, instrumentalness, liveness, speechiness, dan *valence*, serta informasi tambahan seperti durasi lagu, popularitas, dan genre. Adapun rangkuman dari data pada penelitian ini dapat dilihat pada Tabel 3.1.



Tabel 3.1. Rangkuman Dataset

Serial	Track_Id	Artists	Album_Name	...	Track_genre
1	6KwkVtXm8OUp2XffN5k71Y	Hillsong Worship	No Other Name	...	acoustic
2	2dp5I5MJ8bQQHDoFaNRfX	Internal Rot	Grieving Birth	...	acoustic
3	5avw06usmFkFrPjX8NxC40	Zhoobin Askarieh		...	acoustic
...	...	...	...	...	...
113997	6x8ZfSoqDjuNa5SVP5QjvX, CesÃjria			...	World-music
113998	2e6sXL2bYv4bSz6VTdnfLs, Michael			...	World-music
113999	,2hETkH7cOfq mz3LqZDHZf, CesÃjria	Evora, Miss	Perfumado, Barbincor	...	World-music

3.1.2. Prapemrosesan

Praproses data dalam penelitian klasifikasi genre musik pada Spotify menggunakan XGBoost melibatkan beberapa langkah penting untuk memastikan data siap dianalisis secara akurat dan efektif. Langkah pertama adalah pembersihan data dan normalisasi, yang mencakup penghapusan atribut yang tidak relevan seperti nama artis, nama lagu, tahun rilis, dan sebagainya. Selanjutnya, dilakukan konversi data kategorikal menjadi format numerik menggunakan teknik seperti *Label Encoding*. Proses ini juga mencakup penanganan nilai yang hilang dengan menghapus atau mengisi nilai kosong menggunakan metode yang sesuai, serta standarisasi fitur numerik untuk memastikan konsistensi skala antar fitur.

Dengan langkah-langkah praproses ini, dataset menjadi siap untuk diterapkan pada model XGBoost, yang dapat meningkatkan akurasi dalam klasifikasi genre musik berdasarkan fitur audio yang telah diproses sebelumnya.

3.1.3. Analisis Data

Analisis data eksploratori (*Exploratory Data Analysis/EDA*) dilakukan untuk memahami karakteristik data dalam mengidentifikasi pola dan fitur yang relevan dalam proses klasifikasi genre musik. EDA mencakup teknik visualisasi data seperti *histogram* dan *heatmap* korelasi yang digunakan untuk melihat distribusi data dan hubungan antar fitur audio seperti akustik, energi, *loudness*, dan *danceability*. Misalnya, analisis korelasi menunjukkan hubungan negatif yang kuat antara akustik dan energi, serta hubungan positif antara *loudness* dan energi, yang memberikan wawasan penting dalam membangun model klasifikasi genre musik.

3.1.4. Penerapan Algoritma

Dalam penelitian ini, model XGBoost diterapkan untuk mengklasifikasikan genre musik pada platform Spotify berdasarkan fitur audio yang diekstraksi. Implementasi dilakukan menggunakan bahasa pemrograman *Python* dengan pustaka seperti XGBoost. XGBoost dipilih karena kemampuannya dalam menangani data besar dan kompleks, serta efektivitasnya dalam klasifikasi multi-kelas. Evaluasi model dilakukan dengan menggunakan metrik seperti akurasi, presisi, *recall*, dan *F1-score*.

3.1.5. Evaluasi

Tahapan evaluasi bertujuan untuk menilai performa model XGBoost dalam mengklasifikasikan genre musik pada platform Spotify. Evaluasi dilakukan dengan menggunakan dataset pengujian yang telah dipisahkan selama tahap pra-prosesing. Metrik utama yang digunakan untuk mengukur akurasi model adalah akurasi (*accuracy*), yang menghitung persentase prediksi yang benar dari total prediksi yang dilakukan oleh model. Selain itu, digunakan pula metrik evaluasi lainnya seperti presisi (*precision*), *recall*, dan *F1-score* untuk mendapatkan gambaran yang lebih lengkap mengenai performa model dalam mengklasifikasikan genre musik.

Setelah semua metrik evaluasi dihitung, hasilnya dibandingkan untuk menentukan model mana yang memberikan klasifikasi genre musik yang lebih akurat. Hasil evaluasi ini akan digunakan untuk membuat kesimpulan mengenai efektivitas model XGBoost dalam klasifikasi genre musik pada Spotify.

3.2. Perhitungan Manual

Pada penelitian ini, perhitungan manual dilakukan untuk menjelaskan proses yang terjadi dalam algoritma XGBoost yang digunakan dalam klasifikasi genre musik pada Spotify. Perhitungan ini dibagi menjadi beberapa tahapan inti dari algoritma XGBoost, mulai dari inisialisasi prediksi awal, perhitungan *gradient* dan *hessian*, pemilihan fitur terbaik untuk pemisahan (*split*), hingga pembaruan prediksi. Perhitungan manual ini menggunakan data *dummy* atau data contoh yang terdiri dari variabel *artist*, judul lagu, album, popularitas, dan eksplisit. Langkah-langkah perhitungan ini bertujuan untuk memberikan pemahaman yang lebih mendalam mengenai cara kerja XGBoost dalam mengklasifikasikan genre musik secara bertahap.

Berikut adalah contoh perhitungan manual (secara konseptual) klasifikasi genre musik menggunakan algoritma XGBoost dengan data *dummy*. Karena XGBoost melibatkan banyak proses matematis seperti *boosting*, pohon keputusan, dan *gradient descent*, perhitungan manual hanya bisa dilakukan untuk satu iterasi pertama (satu pohon) dan menggunakan data yang sangat sederhana dengan tujuan memprediksi apakah lagu termasuk genre Pop (1) atau Non-Pop (0).

Data Dummy (Fitur)

Data dummy ini digunakan sebagai bahan ilustrasi untuk menjelaskan proses perhitungan secara konseptual.

Tabel 3.2. Data Dummy

ID	Artist	Judul	Album	Popularitas	Eksplisit	Genre (Label)
1	Taylor Swift	Love Story	Fearless	85	0	Pop (1)
2	Eminem	Lose Yourself	Curtain Call	90	1	HipHop (0)
3	Ed Sheeran	Perfect	Divide	88	0	Pop (1)
4	Metallica	One	Justice	70	0	Rock (0)
5	BTS	Dynamite	BE	95	0	Pop (1)

Tabel 3.2 menyajikan contoh data dummy yang digunakan untuk perhitungan manual dalam klasifikasi genre musik menggunakan algoritma XGBoost. Data contoh ini memuat beberapa atribut seperti ID, Artist, Judul lagu, Album, Popularitas, status Eksplisit, dan Label Genre (Pop atau Non-Pop).

1. Encoding Kategori

Proses encoding ini diperlukan agar atribut kategorikal dapat diolah oleh algoritma XGBoost yang hanya menerima input numerik.

Tabel 3.3. Kategori Encoding

Artist	Label
Taylor Swift	0
Eminem	1
Ed Sheeran	2
Metallica	3
BTS	4

Album	Label
Fearless	0
Curtain Call	1
Divide	2
Justice	3
BE	4

Judul	Label
Love Story	0
Lose Yourself	1
Perfect	2
One	3
Dynamite	4

Tabel 3.3 menunjukkan hasil pengkodean (encoding) kategori data pada atribut Artist, Judul lagu, dan Album menjadi bilangan numerik.

2. Dataset Setelah Encoding

Tabel 3.4 memperlihatkan data dummy setelah seluruh atribut kategorikal telah diubah menjadi nilai numerik sesuai dengan hasil encoding.

Tabel 3.4. Setelah Encoding

ID	Artist	Judul	Album	Popularitas	Eksplisit	Genre
1	0	0	0	85	0	1
2	1	1	1	90	1	0
3	2	2	2	88	0	1
4	3	3	3	70	0	0
5	4	4	4	95	0	1

Data ini merupakan input yang digunakan dalam tahap selanjutnya pada perhitungan manual.

3. Inisialisasi Prediksi

Gunakan log-odds = 0:

$$\hat{y}_i^{(0)} = \frac{1}{1 + e^{-0}} = 0.5$$

4. Hitung Gradient dan Hessian

$$g_i = \hat{y}_i - y_i, \quad h_i = \hat{y}_i(1 - \hat{y}_i)$$

Tabel 3.5. Gradient dan Hessian

ID	y_i	$\hat{y}_i$	g_i	h_i
1	1	0.5	-0.5	0.25
2	0	0.5	0.5	0.25
3	1	0.5	-0.5	0.25
4	0	0.5	0.5	0.25
5	1	0.5	-0.5	0.25

Tabel 3.5 menampilkan nilai gradient dan hessian yang dihitung dari prediksi awal dan nilai target pada data dummy. Gradient g_i dan Hessian h_i ini merupakan komponen penting dalam algoritma XGBoost untuk memperbaiki prediksi pada setiap iterasi.

5. Split Berdasarkan Salah Satu Fitur (Misalnya Popularitas < 80)

Left (Pop < 80) → ID 4

Right (Pop ≥ 80) → ID 1,2,3,5

Hitung Gain:

$$\text{Gain} = \frac{1}{2} \left[\frac{(\sum g_L)^2}{\sum h_L + \lambda} + \frac{(\sum g_R)^2}{\sum h_R + \lambda} - \frac{(\sum g_T)^2}{\sum h_T + \lambda} \right]$$

Left: $g = 0.5$, $h = 0.25$

Right: $g = -0.5 + 0.5 - 0.5 = -0.5$, $h = 0.75$

Total: $g = 0, h = 1.0$

$$\lambda = 1$$

$$\text{Gain} = \frac{1}{2} \left[\frac{0.5^2}{0.25+1} + \frac{(-0.5)^2}{0.75+1} \right] = \frac{1}{2} \left[\frac{0.25}{1.25} + \frac{0.25}{1.75} \right] \approx \frac{1}{2} (0.2 + 0.143) = 0.171$$

Hitung Output Leaf

$$w_j = -\frac{\sum g_j}{\sum h_j + \lambda}$$

Left:

$$w_L = -\frac{0.5}{0.25 + 1} = -0.4$$

Right:

$$w_R = -\frac{-0.5}{0.75 + 1} = 0.286$$

Update Prediksi

Learning rate $\eta = 0.1$

$$\text{ID } 1,2,3,5 \rightarrow \hat{y}_i^{(1)} = 0.5 + 0.1 \cdot 0.286 = 0.5286$$

$$\text{ID } 4 \rightarrow \hat{y}_i^{(1)} = 0.5 - 0.1 \cdot 0.4 = 0.46$$

HASIL AKHIR SETELAH 1 POHON:

Prediksi ini menunjukkan nilai probabilitas bahwa masing-masing data termasuk dalam genre Pop. Hasil prediksi mulai menunjukkan kecenderungan yang sesuai, meskipun akurasi belum maksimal.

Tabel 3.6. Hasil Akhir

ID	Genre Asli	Prediksi
1	1	0.5286
2	0	0.5286
3	1	0.5286
4	0	0.4600
5	1	0.5286

Tabel 3.6 memperlihatkan hasil prediksi akhir setelah dilakukan satu iterasi pembelajaran dengan satu pohon keputusan pada data dummy.

Kesimpulan

Prediksi sudah mulai condong ke genre pop (nilai > 0.5)

Akurasi belum maksimal

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma XGBoost mampu memberikan performa klasifikasi yang sangat baik terhadap data musik dari Spotify dengan tiga kategori genre, yaitu *Pop*, *Rock*, dan *Jazz*. Melewati proses praprosesing data, eksplorasi data, serta pemangkasan data terhadap tiga genre utama berhasil menyederhanakan klasifikasi tanpa mengurangi keakuratan model secara signifikan. Eksperimen terhadap nilai *learning rate* (ETA) menunjukkan bahwa nilai 0.08 dan 0.09 menghasilkan performa terbaik dengan akurasi mencapai 88%, serta nilai *precision*, *recall*, dan *f1-score* yang tinggi dan seimbang untuk masing-masing genre. Selain itu, hasil *benchmarking* juga membuktikan bahwa XGBoost secara konsisten mengungguli algoritma pembandingan lainnya seperti *K-Nearest Neighbors*, *Naive Bayes*, dan *Decision Tree* dalam seluruh metrik evaluasi yang digunakan.

5.2. Saran

Penelitian ini masih memiliki ruang untuk pengembangan lebih lanjut. Salah satu saran yang dapat diberikan adalah menambahkan jumlah kelas genre agar hasil klasifikasi menjadi lebih kompleks dan mendekati kondisi nyata. Selain itu, penggunaan teknik optimasi *hyperparameter* yang lebih mendalam seperti *Grid Search* atau *Bayesian Optimization* dapat meningkatkan kinerja model lebih lanjut. Penelitian selanjutnya juga disarankan untuk menggunakan data *real-time* atau data streaming dari API Spotify guna memperluas cakupan dan relevansi hasil.

DAFTAR PUSTAKA

- Álvarez, P., de Quirós, J., & Baldassarri, S. (2023). *RIADA: A machine-learning based infrastructure for recognising the emotions of spotify songs*.
- Barate, A., Ludovico, L. A., &. (2022). A web platform to extract and investigate music genre labels in spotify. *Proceedings of the Sound and Music Computing Conferences*, 369–376.
- Barranco-Chamorro, I., & Carrillo-García, R. M. (2021). Techniques to deal with off-diagonal elements in confusion matrices. *Mathematics*, 9(24).
<https://doi.org/10.3390/math9243233>
- Castillo, J. R., & Flores, M. J. (2021). Web-based music genre classification for timeline song visualization and analysis. *IEEE Access*, 9, 18801–18816.
- Cendani, L. M., & Wibowo, A. (2022). Perbandingan Metode ensemble learning pada klasifikasi penyakit diabetes. *Jurnal Masyarakat Informatika*, 13(1), 33–44.
- Ceylan, H. C., Hardalaç, N., Kara, A. C., & others. (2021). Automatic music genre classification and its relation with music education. *World Journal of Education*, 11(2), 36–45.
- Chaudhury, M., Karami, A., & Ghazanfar, M. A. (2022). Large-scale music genre analysis and classification using machine learning with apache spark. *Electronics*, 11(16), 2567.
- De Nadai, M., Fabbri, F., Gigioli, P., Wang, A., Li, A., Silvestri, F., Kim, L., Lin, S., Radosavljevic, V., Ghael, S., & others. (2024). Personalized audiobook recommendations at spotify through graph neural networks. *Companion*

Proceedings of the ACM Web Conference 2024, 403–412.

Demir, A. E. (2022). Music Listening Practices and Spotify App in the Context of Everyday Life Sociology. *Demir, AE (2022). Gündelik Hayat Sosyolojisi Ba{\u{g}}lam{\i}nda Müzik Dinleme Pratikleri ve Spotify Uygulamas{\i}. Haliç Üniversitesi Sosyal Bilimler Dergisi*, 5(1), 163–188.

Deruty, E., Grachten, M., Lattner, S., Nistal, J., & Aouameur, C. (2022). On the development and practice of ai technology for contemporary popular music production. *Transactions of the International Society for Music Information Retrieval*, 5(1).

Effectiveness, T., & Regression, L. (2024). *Efektivitas Algoritma Random Forest , XGBoost , dan Logistic Regression dalam Prediksi Penyakit Paru-paru*. 23(4), 909–922.

EL Fadel, N. (2025). Facial Recognition Algorithms: A Systematic Literature Review. *Journal of Imaging*, 11(2), 58.

Epstein, E., Nallapareddy, N., & Ray, S. (2023). On the Relationship between Feature Selection Metrics and Accuracy. *Entropy*, 25(12), 1–22.
<https://doi.org/10.3390/e25121646>

Fadli, M., & Saputra, R. A. (2023). Klasifikasi Dan Evaluasi Performa Model Random Forest Untuk Prediksi Stroke. *JT: Jurnal Teknik*, 12(02), 72–80.
<http://jurnal.umt.ac.id/index.php/jt/index>

Ginting, J., Ginting, R., & Hartono, H. (2022). Deteksi Dan Prediksi Penyakit Diabetes Melitus Tipe 2 Menggunakan Machine Learning (Scooping Review). *Jurnal Keperawatan Priority*, 5(2), 93–105.
<https://doi.org/10.34012/jukep.v5i2.2671>

- Hendri, D., Nadha, D., Basri, F. K., Wajdi, M. F., & Nadhirah, N. (2024). Comparison of Decision Tree Algorithm, Naive Bayes, K-Nearest Neighbors on Spotify Music Genre. *IJATIS: Indonesian Journal of Applied Technology and Innovation Science*, 1(1), 47–53.
- Hesmondhalgh, D., Valverde, R. C., Kaye, D. B. V., & Li, Z. (2024). Critically Analyzing Platform Interfaces: How Music-Streaming Platforms Frame Musical Experience. *International Journal of Communication*, 18, 22.
- Indonesia, P. B. (2025). *MACHINE LEARNING UNTUK PEMULA (Konsep dan Implementasi)* (Issue April).
- Ismanto, E., & Novalia, M. (2021). Komparasi Kinerja Algoritma C4.5, Random Forest, dan Gradient Boosting untuk Klasifikasi Komoditas. *Techno.Com*, 20(3), 400–410. <https://doi.org/10.33633/tc.v20i3.4576>
- Kiriakidou, N., Livieris, I. E., & Diou, C. (2024). C-XGBoost: A tree boosting model for causal effect estimation. *IFIP International Conference on Artificial Intelligence Applications and Innovations*, 58–70.
- Kostrzewa, D., Kaminski, P., & Brzeski, R. (2021). Music genre classification: looking for the perfect network. *International Conference on Computational Science*, 55–67.
- Kufel, J., Bargieł-Łkaczek, K., Kocot, S., Koźlik, M., Bartnikowska, W., Janik Michałand Czogalik, Ł., Dudek, P., Magiera, M., Lis, A., & others. (2023). What is machine learning, artificial neural networks and deep learning?— Examples of practical applications in medicine. *Diagnostics*, 13(15), 2582.
- Lindholm, J. (2022). *Hit song classification with spotify audio features of Billboard list songs*.

- LUTFIANI, S., Saragih, T. H., Abadi, F., Faisal, M. R., & Kartini, D. (2023). Perbandingan Metode Extreme Gradient Boosting Dan Metode Decision Tree Untuk Klasifikasi Genre Musik. *Jurnal Informatika Polinema*, 9(4), 373–382. <https://doi.org/10.33795/jip.v9i4.1319>
- Maesaroh, S., Mubarak, R., Hakim, L., Yunianto, I., Mutmainah, S., Santoso, H., Utami, W. S., Nugroho, A. Y., Soleh, O., Oktavian, R., & others. (2024). *Pembelajaran Mesin dan Kecerdasan Buatan: Teori dan Aplikasi Praktis*. Sada Kurnia Pustaka.
- Meléndez, R., Ptaszynski, M., & Masui, F. (2024). Comparative Investigation of Traditional Machine-Learning Models and Transformer Models for Phishing Email Detection. *Electronics (Switzerland)*, 13(24). <https://doi.org/10.3390/electronics13244877>
- Moshawrab, M., Adda, M., Bouzouane, A., Ibrahim, H., & Raad, A. (2023). Reviewing federated learning aggregation algorithms; strategies, contributions, limitations and future perspectives. *Electronics*, 12(10), 2287.
- Mousavi, A., Pourdarbani, R., Sabzi, S., Sotoudeh, D., Moradzadeh, M., García-Mateos, G., Kasaei, S., & Rohban, M. H. (2024). Classification of Healthy and Frozen Pomegranates Using Hyperspectral Imaging and Deep Learning. *Horticulturae*, 10(1). <https://doi.org/10.3390/horticulturae10010043>
- Musyarofah, U. L., Alima, S. N., & Kartika, D. S. Y. (2022). Klasifikasi Top 50 Spotify Tahun 2010-2019 Menggunakan Metode K-Means Clustering. *Prosiding Seminar Nasional Teknologi Dan Sistem Informasi*, 2(1), 215–220.
- Naufal, M. F. (2021). Analisis Perbandingan Algoritma Svm, Knn, Dan Cnn untuk Klasifikasi Citra Cuaca. *Jurnal Teknologi Informasi Dan Ilmu*

Komputer, 8(2), 311–317.

Navisa, S., Hakim, L., & Nabilah, A. (2021). Komparasi Algoritma Klasifikasi Genre Musik pada Spotify Menggunakan CRISP-DM. *Jurnal Sistem Cerdas*, 4(2), 114–125.

Noh, B., Youm, C., Goh, E., Lee, M., Park, H., Jeon, H., & Kim, O. Y. (2021). XGBoost based machine learning approach to predict the risk of fall in older adults using gait outcomes. *Scientific Reports*, 11(1), 12183.

Nurhalizah, R. S., Ardianto, R., & Purwono, P. (2024). Analisis Supervised dan Unsupervised Learning pada Machine Learning: Systematic Literature Review. *Jurnal Ilmu Komputer Dan Informatika*, 4(1), 61–72.

Nurussakinah, N., & Faisal, M. (2023). Klasifikasi Penyakit Diabetes Menggunakan Algoritma Decision Tree. *Jurnal Informatika*, 10(2), 143–149. <https://doi.org/10.31294/inf.v10i2.15989>

Oliveira, G. P., Barbosa, G. R. G., Melo, B. C., Botelho, J. E., Silva, M. O., Seufitelli, D. B., & Moro, M. M. (2022). Musical success in the United States and Brazil: Novel datasets and temporal analyses. *Journal of Information and Data Management*, 13(1).

Pratiwi, E. Y., Haq, A. Z., & Daufa, Z. D. (2025). AI DALAM MANAJEMEN RISIKO UNTUK MEMBANGUN KEPUTUSAN KEUANGAN YANG LEBIH BAIK. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(1), 999–1004.

Qin, R. (2022). The construction of corporate financial management risk model based on XGBoost algorithm. *Journal of Mathematics*, 2022(1), 2043369.

Randi Farmana Putra, M. S., Iqbal Ramadhani Mukhlis, S.Kom., M.Kom Aulia

- Iefan Datya, ST., M., Sio Jurnalis Pipin, S.Kom., M.Kom Felix Reba, S.Si., M. S., Muhammad Al-Husaini, S.T., M. K., & Samuel Aleksander Mandowen, S.Si., M.IT Neni Nur Laili Ersela Zain, S.Si., M.Mat Loso Judijanto, S.Si., M.M., M. S. (2016). *ALGORITMA PEMBELAJARAN MESIN (Dasar, Teknik, dan Aplikasi)* (Vol. 19, Issue 5).
- Sabiri, B., Khtira, A., El Asri, B., & Rhanoui, M. (2024). Investigating Contrastive Pair Learning's Frontiers in Supervised, Semisupervised, and Self-Supervised Learning. *Journal of Imaging*, 10(8), 196.
- Safitri, A. R., & Arifin Zaenal. (2023). Analisis Akurasi Model-model Prediksi Kebangkrutan: Studi pada Perusahaan Go Public yang Mengalami Financial Distress. *Selekta Manajemen : Jurnal Mahasiswa Bisnis & Manajemen*, 02(03), 323–344.
- Saputro, M. B., & Alamsyah, A. (2024). Comparison Of Naive Bayes Classifier And K-Nearest Neighbor Algorithms With Information Gain And Adaptive Boosting For Sentiment Analysis Of Spotify App Reviews. *Recursive Journal of Informatics*, 2(1), 37–44.
- Sari, M., & Mahalisa, G. (2024). Naive Bayes Classifier Untuk Deteksi Email Spam. *Technologia: Jurnal Ilmiah*, 15(4), 675–680.
- Septiana Rizky, P., Haiban Hirzi, R., & Hidayaturrohman, U. (2022). Perbandingan Metode LightGBM dan XGBoost dalam Menangani Data dengan Kelas Tidak Seimbang. *J Statistika: Jurnal Ilmiah Teori Dan Aplikasi Statistika*, 15(2), 228–236.
<https://doi.org/10.36456/jstat.vol15.no2.a5548>
- Setiawan, M. B., & Rahmatulloh, A. (2025). *Analisis Perbandingan Model*

Random Forest dan XGBoost dalam Memprediksi Turnover Karyawan.

15(2), 393–400.

Simanjuntak, T. I., Muhathir, M., Fadlisyah, F., & Safira, I. (2023). Performance Analysis of Naive Bayes Variation Method in Spice Image Classification Using Histogram of Gradient Oriented (HOG) Feature Extraction. *Journal of Informatics and Telecommunication Engineering*, 7(1), 282–291.

<https://doi.org/10.31289/jite.v7i1.7957>

Strelcenia, E., & Prakoonwit, S. (2023). Improving classification performance in credit card fraud detection by using new data augmentation. *AI*, 4(1).

Suryana, S. E., Warsito, B., & Suparti, S. (2021). Penerapan Gradient Boosting Dengan Hyperopt Untuk Memprediksi Keberhasilan Telemarketing Bank. *Jurnal Gaussian*, 10(4), 617–623.

<https://doi.org/10.14710/j.gauss.v10i4.31335>

Tahiyat, A. N., Maulana, B., Saputra, A. E., Efrizoni, L., & Rahmaddeni, R. (2025). Klasterisasi Lagu Populer dan Eksplorasi Subgenre Spotify 2024 dengan K-Medoids. *Jurnal Informatika Dan Teknologi Komputer (JITEK)*, 5(1), 34–48.

Tannady, H., Fernandes, J., Honni, A., & Lee, S. (2024). Analisis Big Data Spotify dengan Metode Data Mining Analysis of Spotify Big Data with Data Mining Method. 7(2), 52–59.

Tantyoko, H., Sari, D. K., & Wijaya, A. R. (2023). Prediksi Potensial Gempa Bumi Indonesia Menggunakan Metode Random Forest Dan Feature Selection. *IDEALIS : InDonEsiA Journal Information System*, 6(2), 83–89.

<https://doi.org/10.36080/idealis.v6i2.3036>

- Tran, N. T., Tran, T. T. G., Nguyen, T. A., & Lam, M. B. (2023). A new grid search algorithm based on XGBoost model for load forecasting. *Bulletin of Electrical Engineering and Informatics*, 12(4), 1857–1866.
- Tuning, H. (2025). *Evaluasi Performa XGBoost dengan Oversampling dan Hyperparameter Tuning untuk Prediksi Alzheimer*. 24(1), 1–12.
<https://doi.org/10.62411/tc.v24i1.12057>
- Turaina, R., & Saputra, R. (2024). *Optimalisasi Deteksi Malware pada Platform Android dengan Pendekatan Ensemble Machine Learning*. 7(3), 394–403.
- Wairata, C. R., Swedia, E. R., & Cahyanti, M. (2021). Pengklasifikasian Genre Musik Indonesia Menggunakan Convolutional Neural Network. *Sebatik*, 25(1), 255–261. <https://doi.org/10.46984/sebatik.v25i1.1286>
- Zhang, C., Wang, D., Wang, L., Guan, L., Yang, H., Zhang, Z., Chen, X., & Zhang, M. (2021). Cause-aware failure detection using an interpretable XGBoost for optical networks. *Optics Express*, 29(20), 31974–31992.
- Zhang, Y., & Chen, L. (2021). A study on forecasting the default risk of bond based on xgboost algorithm and over-sampling method. *Theoretical Economics Letters*, 11(2), 258–267.

LAMPIRAN

1. Kode Program

```
import pandas as pd

from google.colab import drive

drive.mount('/content/drive')

import pandas as pd

df = pd.read_csv('/content/drive/MyDrive/Colab
Notebooks/train.csv')

df.head()

df.info()

#Seleksi Genre

genre_counts = df['track_genre'].value_counts()

genre_counts

# prompt: select several rows data with specific
'track_genre'

target_genres = ['pop', 'rock', 'jazz']

selected_df = df[df['track_genre'].isin(target_genres)]

selected_df.info()


#Transformasi Data Kategorikal menjadi Numerik

from sklearn.preprocessing import LabelEncoder

le = LabelEncoder()

selected_df['track_genre'] =
le.fit_transform(selected_df['track_genre'])

selected_df.head()
```

```
#Penghapusan atribut yang tidak digunakan

df1 = selected_df.drop(['Unnamed: 0', 'track_id', 'artists',
'album_name', 'track_name'], axis=1)

df1.info()

df1.describe()

#Proses Normalisasi Data

# prompt: normalize df1 using minmax

from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler()

df1_scaled = scaler.fit_transform(df1)

df1_normalized = pd.DataFrame(df1_scaled,
columns=df1.columns)

df1_normalized.head()

df1_normalized.info()

#Heatmap Korelasi

import matplotlib.pyplot as plt

import seaborn as sns

# Calculate the correlation matrix

correlation_matrix = df1_normalized.corr()

# Plot the correlation matrix as a heatmap

plt.figure(figsize=(10, 8))

sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm',
fmt=".2f")

plt.title('Correlation Matrix of Normalized DataFrame')

plt.show()
```

```
#Histogram

df1.hist(figsize=(15, 10))

plt.tight_layout()

plt.show()

#Implementasi Algoritma XGBoost

#Pembagian X dan Y

X = df1_normalized.drop('track_genre', axis=1)
y = selected_df['track_genre']

#Instalasi XGBoost

!pip install xgboost

#Penerapan Algoritma XGBoost

import xgboost as xgb

from sklearn.model_selection import train_test_split

from sklearn.metrics import accuracy_score,
classification_report

# Split data into training and testing sets

X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Initialize XGBoost Classifier

# For classification problems, 'objective' should be
'binary:logistic' or 'multi:softmax'/'multi:softprob'

# depending on the number of classes. Since 'track_genre' is
multi-class after encoding,

# 'multi:softmax' or 'multi:softprob' is appropriate. We'll
use 'multi:softmax'.
```



```
# We also need 'num_class' for multi-class problems.

num_classes = 3 # Get the number of unique classes

model = xgb.XGBClassifier(objective='multi:softmax',
num_class=num_classes, eval_metric='merror', eta=0.09,
use_label_encoder=False)

# Train the model

model.fit(X_train, y_train)

# Make predictions on the test set

y_pred = model.predict(X_test)

# Evaluate the model

accuracy = accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy:.4f}")

# Print classification report
print("Classification Report:")
print(classification_report(y_test, y_pred))

# prompt: confusion matrix plot
from sklearn.metrics import confusion_matrix

# Compute the confusion matrix

cm = confusion_matrix(y_test, y_pred)

# Plot the confusion matrix

plt.figure(figsize=(10, 8))

sns.heatmap(cm, annot=True, fmt='d', cmap='Blues',
xticklabels=le.classes_, yticklabels=le.classes_,
annot_kws={"size": 15})

plt.xlabel('Predicted Label', fontsize=20)

plt.ylabel('True Label', fontsize=20)
```

```
plt.title('Confusion Matrix', fontsize=20)

plt.show()

# prompt: knn implementation

from sklearn.neighbors import KNeighborsClassifier

# Pembagian X dan Y (already done in the preceding code)

# X = df1_normalized.drop('track_genre', axis=1)

# y = df['track_genre']

# Split data into training and testing sets (already done)

# X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Initialize KNN Classifier

# We can start with a small number of neighbors, e.g., k=5

knn_model = KNeighborsClassifier(n_neighbors=5)

# Train the KNN model

knn_model.fit(X_train, y_train)

# Make predictions on the test set using KNN

y_pred_knn = knn_model.predict(X_test)

# Evaluate the KNN model

accuracy_knn = accuracy_score(y_test, y_pred_knn)

print(f"KNN Accuracy: {accuracy_knn:.4f}")

# Print KNN classification report

print("KNN Classification Report:")

print(classification_report(y_test, y_pred_knn))
```

```
# prompt: confusion matrix knn

# Compute the confusion matrix for KNN

cm_knn = confusion_matrix(y_test, y_pred_knn)


# Plot the confusion matrix for KNN

plt.figure(figsize=(10, 8))

sns.heatmap(cm_knn, annot=True, fmt='d', cmap='Blues',
            xticklabels=le.classes_, yticklabels=le.classes_,
            annot_kws={"size": 15})

plt.xlabel('Predicted Label', fontsize=20)
plt.ylabel('True Label', fontsize=20)
plt.title('Confusion Matrix for KNN', fontsize=20)
plt.show()


# prompt: implement naive bayes

from sklearn.naive_bayes import GaussianNB


# Initialize Naive Bayes Classifier (GaussianNB is suitable
for continuous features after normalization)

nb_model = GaussianNB()


# Train the Naive Bayes model

nb_model.fit(X_train, y_train)


# Make predictions on the test set using Naive Bayes

y_pred_nb = nb_model.predict(X_test)
```

```

# Evaluate the Naive Bayes model

accuracy_nb = accuracy_score(y_test, y_pred_nb)

print(f"Naive Bayes Accuracy: {accuracy_nb:.4f}")


# Print Naive Bayes classification report

print("Naive Bayes Classification Report:")

print(classification_report(y_test, y_pred_nb))


# prompt: confusion matrix naive bayes

# Compute the confusion matrix for Naive Bayes

cm_nb = confusion_matrix(y_test, y_pred_nb)


# Plot the confusion matrix for Naive Bayes

plt.figure(figsize=(10, 8))

sns.heatmap(cm_nb, annot=True, fmt='d', cmap='Blues',
            xticklabels=le.classes_, yticklabels=le.classes_,
            annot_kws={"size": 15})

plt.xlabel('Predicted Label', fontsize=20)

plt.ylabel('True Label', fontsize=20)

plt.title('Confusion Matrix for Naive Bayes', fontsize=20)

plt.show()


# prompt: implement decision tree

from sklearn.tree import DecisionTreeClassifier

```

```
# Initialize Decision Tree Classifier

# You can tune parameters like criterion ('gini' or
'entropy'), max_depth, min_samples_split, etc.

dt_model = DecisionTreeClassifier(random_state=42)

# Train the Decision Tree model

dt_model.fit(X_train, y_train)

# Make predictions on the test set using Decision Tree

y_pred_dt = dt_model.predict(X_test)

# Evaluate the Decision Tree model

accuracy_dt = accuracy_score(y_test, y_pred_dt)
print(f"Decision Tree Accuracy: {accuracy_dt:.4f}")

# Print Decision Tree classification report
print("Decision Tree Classification Report:")
print(classification_report(y_test, y_pred_dt))

# prompt: confusion matrix decision tree

# Compute the confusion matrix for Decision Tree

cm_dt = confusion_matrix(y_test, y_pred_dt)

# Plot the confusion matrix for Decision Tree

plt.figure(figsize=(10, 8))
```



```
sns.heatmap(cm_dt, annot=True, fmt='d', cmap='Blues',  
xticklabels=le.classes_, yticklabels=le.classes_,  
annot_kws={"size": 15})  
  
plt.xlabel('Predicted Label', fontsize=20)  
  
plt.ylabel('True Label', fontsize=20)  
  
plt.title('Confusion Matrix for Decision Tree', fontsize=20)  
  
plt.show()
```



PAPER NAME

BIMO ZULFI ANANDA PUTRA_Klasifikasi
Genre Musik Pada Spotify Dengan Meng
gunakan XGBoost_EAeMkuMZWDfUUKSI
FBGNswL2ZCR2UM8gZfsynmFC.pdf

AUTHOR

BIMO ZULFI ANANDA PUTRA



WORD COUNT

10759 Words

CHARACTER COUNT

66346 Characters

PAGE COUNT

65 Pages

FILE SIZE

1.2MB

SUBMISSION DATE

Dec 10, 2025 2:59 PM GMT+7

REPORT DATE

Dec 10, 2025 3:01 PM GMT+7

● 10% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 6% Internet database
- 2% Publications database
- Crossref database
- Crossref Posted Content database
- 6% Submitted Works database

● Excluded from Similarity Report

- Bibliographic material
- Cited material
- Abstract
- Small Matches (Less than 15 words)

Summary



UNIVERSITAS MEDAN AREA FAKULTAS TEKNIK

Kampus I : Jalan Kualanaram Nomor 1 Medan Estate / Jalan Gedung PDSI, Medan 20223
Kampus II : Jalan Sei Serayu Nomor 70 A / Jalan Setia Budi Nomor 79 D, Medan 20112 Telepon : (061) 8225602, 8201994
Fax : (061) 8226331 HP : (081) 607 259 website: www.uma.ac.id Email : umv@medanarea.uma.ac.id

Nomor : 1040/FT/01.10/V/2025
Lampiran : -
Hal : Pembimbing Tugas Akhir

23 Mei 2025

Yth. Pembimbing Tugas Akhir
ANDRE HASUDUNGAN LBS S.TI, M.Sc (Sebagai Pembimbing)
di Tempat

Dengan hormat, sehubungan telah dipenuhinya persyaratan untuk memperoleh Tugas Akhir dari mahasiswa atas :

Nama : BIMO ZULFI ANANDA PUTRA
NIM : 218160031
Jurusan : TEKNIK INFORMATIKA

Maka dengan hormat kami mengharapkan kesediaan saudara :

ANDRE HASUDUNGAN LBS S.TI, M.Sc (Sebagai Pembimbing)

Adapun Tugas Akhir Skripsi berjudul :

KLASIFIKASI GENRE MUSIK PADA SPOTIFY DENGAN MENGGUNAKAN XGBOOST

SK Pembimbing ini berlaku selama enam bulan terhitung sejak SK ini diterbitkan. Jika proses pembimbing melebihi batas waktu yang telah ditetapkan, SK ini dapat ditinjau ulang.

Demikian kami sampaikan, atas kesediaan saudara diucapkan terima kasih.

Dekan,



Dr Eng. Supriatno.ST, MT.





UNIVERSITAS MEDAN AREA

FAKULTAS TEKNIK

Kampus I : Jalan Kolam Nomor 1 Medan Estate (061) 7360158, Medan, 20223
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A (061) 4243294, Medan, 20122
Website: www.teknik.uma.ac.id E-mail: univ_medanarea@uma.ac.id

Nomor : 190/FT.6/01.10/VII/2025
Lamp : -
Hal : Penelitian Dan Pengambilan Data Tugas Akhir

1 Juli 2025

Yth. Wakil Rektor Bidang Mutu Sumber Daya Manusia dan Perekonomian
Jln. Kolam No.1
Di
Medan

Dengan hormat, kami mohon kesediaan bapak kiranya berkenan untuk memberikan izin dan kesempatan kepada mahasiswa kami tersebut dibawah ini :

NO	NAMA	NPM	PRODI
1	Bimo Zulfi Ananda Putra	218160031	Teknik Informatika

Untuk melaksanakan Penelitian dan Pengambilan Data Tugas Akhir di **Laboratorium Komputer Program Studi Teknik Informatika Fakultas Teknik Universitas Medan Area**.

Perlu kami jelaskan bahwa Pengambilan Data tersebut adalah semata-mata untuk tujuan Ilmiah dan Skripsi, yang merupakan salah satu syarat bagi mahasiswa tersebut untuk mengikuti ujian sarjana pada Fakultas Teknik Universitas Medan Area dan tidak untuk dipublikasikan, dengan judul :

Klasifikasi Genre Musik pada Spotify dengan Menggunakan XGBOOST.

Mohon kiranya tanggal Surat Izin Pengambilan Data Tugas Akhir agar disesuaikan dengan tanggal Terbitnya Surat ini.

Atas perhatian dan kerja sama yang baik diucapkan terima kasih.

Dekan

Dr. Eng. Supriatno, ST, MT

Tembusan :
1. Ka. BPMPP
2. Mahasiswa
3. File



UNIVERSITAS MEDAN AREA

Kampus I : Jalan Kolam Nomor 1 Medan Estate ☎ (061) 7360168, Medan 20223
Kampus II : Jalan Setia Budi Nomor 79 / Jalan Sei Serayu Nomor 70 A ☎ (061) 42402994, Medan 20122
Website: www.uma.ac.id E-Mail: univ_medanarea@uma.ac.id

SURAT KETERANGAN Nomor : 1336/UMA/B/01.7/VIII/2025

Rektor Universitas Medan Area dengan ini menerangkan bahwa :

Nama : Bimo Zulfi Ananda Putra
No.Pokok Mahasiswa : 218160031
Program Studi : Teknik Informatika

Benar telah selesai Pengambilan Data di Laboratorium Komputer Program Studi Teknik Informatika Fakultas Teknik Universitas Medan Area dan jangka waktu penelitian dengan Judul Skripsi **"Klasifikasi Genre Musik pada Spotify dengan Menggunakan XGBOOST"**

Dan kami harapkan Data tersebut kiranya dapat membantu yang bersangkutan dalam penyusunan Skripsi dan dapat bermanfaat bagi mahasiswa khususnya Fakultas Teknik Universitas Medan Area.

Demikian surat ini diterbitkan untuk dapat digunakan seperlunya

Medan, 06 Agustus 2025
an R e k t o r
Wakil Rektor Bidang Mutu SDM &
Perekonomian,


Dr. Dedi Sahputra, S.Sos, MA

CC :
- Arsip

