

Final Project
**Klasifikasi Audio Musik Berdasarkan Penggunaan Autotune Menggunakan Feature
Extraction dan Machine Learning**



Disusun Oleh:
Kelompok 9

Khansa Mutiara Kheeda	(162112133110)
Shiba Salsabilla	(164221078)
Nazhifah Firyal Jasmine	(164221083)
Aysha Milarta Shafiyalina	(164221122)

PROGRAM STUDI S1
TEKNOLOGI SAINS DATA
FAKULTAS TEKNOLOGI MAJU DAN MULTIDISIPLIN
UNIVERSITAS AIRLANGGA
SURABAYA
2024

DAFTAR ISI

BAB I.....	4
PENDAHULUAN.....	4
1.1 Latar Belakang.....	4
1.2 Rumusan Masalah.....	5
1.3 Tujuan.....	5
1.4 Manfaat.....	5
BAB II.....	6
TINJAUAN PUSTAKA.....	6
2.1 Audio Suara.....	6
2.2 Musik.....	6
2.3 Autotune.....	6
2.4 Feature Extraction.....	6
2.4.1 MFCC.....	7
2.4.2 Spectral Centroid.....	7
2.4.3 Chroma.....	7
2.5 Klasifikasi dengan Machine Learning.....	7
2.5.1 Random Forest.....	7
2.5.2 Support Vector Machine.....	7
2.5.3 Multi Layer Perceptron.....	8
BAB III.....	9
METODOLOGI.....	9
3.1 Sumber Data.....	9
3.2 Metodologi.....	9
3.2.1 Pengumpulan data.....	9
3.2.2 Eksploratory Data Analysis (EDA).....	10
3.2.3 Preprocessing.....	10
3.2.4 Ekstraksi fitur audio.....	10
3.2.5 Klasifikasi Model.....	10
3.2.6 Evaluasi Model.....	10
BAB IV	
HASIL DAN PEMBAHASAN.....	11
4.1 Eksplorasi Data Analisis.....	11
4.1.1 Data Information.....	11
4.1.2 Bar Chart Kategori.....	11
4.1.4 Distribusi Durasi Audio.....	12
4.1.5 Visualisasi Fitur	
Visualisasi di atas menunjukkan analisis audio untuk kelas "Autotune" menggunakan	
file dengan indeks ke-0 atau file pertama.....	12
4.2 Preprocessing.....	14
4.2.1 Data Augmentation.....	14

4.2.3 SMOTE.....	15
4.3 Evaluasi Kinerja Feature Extraction.....	15
4.3.1 Random Forest.....	15
4.3.2 Support Vector Machine.....	16
4.3.3 Multi-Layer Perceptron.....	17
BAB V	
KESIMPULAN.....	18
5.1 Kesimpulan.....	18
5.2 Saran.....	18
DAFTAR PUSTAKA.....	19
Lampiran.....	21

Daftar Tabel

Tabel 3.1 Variabel Penelitian.....	9
Tabel 4.1.1 Data Information.....	11
Tabel 4.3.1 Random Forest.....	15
Tabel 4.3.2 Support Vector Machine.....	16
Tabel 4.3.3 Multi-Layer Perceptron.....	17

Daftar Gambar

Gambar 1. Alur Penelitian.....	9
Gambar 2. Bar Chart Distribusi Kategori Dataset.....	11
Gambar 3. Distribusi Durasi Audio.....	12
Gambar 4. Visualisasi Fitur pada Autotune.....	13
Gambar 5. Visualisasi Fitur pada no autotune.....	14
Gambar 6. Data Augmentation.....	14
Gambar 7. Kelas Sebelum dan Sesudah SMOTE.....	15

BAB I PENDAHULUAN

1.1 Latar Belakang

Industri musik telah berkembang pesat seiring dengan kemajuan teknologi. Teknologi dalam industri musik telah lama diimplementasikan sejak tahun 1950-an ketika ilmuwan komputer bereksperimen dengan algoritma untuk menciptakan musik (Viznyak, 2024). Perkembangan industri musik yang memanfaatkan teknologi salah satunya adalah *autotune*. Autotune terdiri dari mengedit sebuah karya musik untuk mengubah atau memperbaiki nada vokal seorang penyanyi. Ketika seorang penyanyi menyanyikan nada yang sedikit fals, autotune secara otomatis dapat menyesuaikan nada tersebut agar selaras dan membantu memperhalus ketidaksempurnaan dalam penampilan vokal dan menciptakan hasil akhir yang lebih rapi (Gohari dkk., 2024).

Penggunaan autotune sering dikritik karena membuat vokal terdengar seperti robot dan memperbaiki kualitas vokal penyanyi yang kurang berbakat. Vokalis Death Cab For Cutie, Ben Gibbard, menyatakan di Grammy Awards 2009 bahwa dalam 10 tahun terakhir, banyak musisi yang terpengaruh oleh manipulasi digital terhadap suara manusia dan mengajak para penyanyi untuk benar-benar menyanyi dan terdengar seperti manusia (Fox et al., 2024).

Untuk mengidentifikasi adanya manipulasi suara yang dilakukan oleh penyanyi, dilakukan klasifikasi suara yang menggunakan autotune dan tidak menggunakan autotune. Klasifikasi merupakan salah satu metode yang digunakan untuk menemukan sebuah model atau fungsi yang menjelaskan dan mencirikan konsep atau kelas data, untuk kepentingan tertentu. Beragam teknik klasifikasi audio telah dikembangkan yang bertujuan untuk memperbaiki dan meningkatkan efisiensi akurasi, pengenalan pola, pemrosesan sinyal, ekstraksi dan tingkat pengenalan untuk menghasilkan klasifikasi yang akurat (Mustaqim dkk., 2019). Ada berbagai teknik yang dapat digunakan untuk mengekstraksi informasi yang ada di dalam sebuah audio salah satunya dengan menggunakan feature extraction. Feature extraction merupakan proses mengubah segmen audio menjadi bentuk numerik yang mana merepresentasikan karakteristik dari audio yang bersangkutan. Feature extraction pada umumnya berupa beberapa bentuk representasi dari waktu dan frekuensi (Laksana & Sulianta, 2017). Berdasarkan Librosa Documentation, terdapat beberapa jenis *feature extraction* seperti *spectral features*, *rhythm features*, *feature manipulation*, dan *feature inversion*. Beberapa teknik menghasilkan pola dari fitur-fitur tersebut dan menggunakannya untuk klasifikasi berdasarkan tingkat korelasi. Teknik lainnya menggunakan nilai numerik dari fitur-fitur tersebut yang digabungkan dengan metode klasifikasi statistik.

Hasil dari proses feature extraction berupa representasi numerik yang menggambarkan karakteristik audio selanjutnya akan diolah menggunakan model machine learning. Machine learning sendiri didefinisikan sebagai penerapan komputer dan algoritma matematika yang dirancang untuk mempelajari pola dari data dan menghasilkan prediksi untuk masa depan yang melibatkan dua tahapan utama, yaitu pelatihan (training) untuk membangun model dan pengujian (testing) untuk mengevaluasi kinerjanya (Roihan dkk., 2020). Algoritma machine learning yang akan digunakan meliputi Random Forest, Support Vector Machine, dan

Multi-Layer Perceptron. Algoritma tersebut dipilih berdasarkan penelitian Atahan dkk pada tahun 2021 yang berhasil mengimplementasikan model serupa untuk klasifikasi genre musik dengan hasil yang optimal. Penelitian ini bertujuan untuk mengevaluasi kualitas suara penyanyi melalui klasifikasi audio yang menggunakan autotune dan tidak menggunakan autotune, serta memberikan wawasan kepada pendengar musik mengenai pengaruh teknologi terhadap kualitas vokal penyanyi dengan memanfaatkan algoritma machine learning.

1.2 Rumusan Masalah

Adapun rumusan masalah dalam penelitian ini, yaitu:

1. Bagaimana karakteristik fitur spektral yang membedakan audio musik dengan *autotune* dan tanpa *autotune*?
2. Bagaimana membangun model klasifikasi yang efektif untuk mengidentifikasi audio musik dengan *autotune* dan tanpa *autotune*?
3. Bagaimana performa model klasifikasi yang digunakan dapat membedakan audio musik yang menggunakan *autotune*?

1.3 Tujuan

Adapun tujuan dalam penelitian ini, yaitu:

1. Mengekstraksi dan menganalisis fitur-fitur spektral yang dapat membedakan musik dengan dan tanpa *autotune*.
2. Mengembangkan model klasifikasi yang mampu mengidentifikasi adanya *autotune* dalam rekaman audio.
3. Menguji performa berbagai algoritma *machine learning* dalam mengklasifikasi audio berdasarkan penggunaan *autotune*.

1.4 Manfaat

Adapun manfaat dilakukannya penelitian ini, yaitu:

1. Menyediakan metode untuk mendeteksi penggunaan *autotune* yang dapat membantu industri musik mengidentifikasi adanya manipulasi suara.
2. Mengembangkan algoritma klasifikasi audio.
3. Berkontribusi dalam bidang klasifikasi audio berbasis *machine learning*.

BAB II

TINJAUAN PUSTAKA

2.1 Audio Suara

Suara dapat ditemukan di berbagai tempat dalam aktivitas di luar ruangan maupun di dalam ruangan yang dihasilkan oleh kegiatan manusia maupun non-manusia. Suara yang dapat didengar dikategorikan menjadi ucapan, musik, dan suara lingkungan. Ucapan dihasilkan oleh manusia dengan menggunakan kombinasi berbagai organ seperti paru-paru, mulut, hidung, dan perut. Suara yang dihasilkan memiliki frekuensi dari 100 Hz hingga 17kHz. Suara musik dihasilkan oleh alat musik atau manusia untuk menciptakan harmoni dan ekspresi emosi. Rentang frekuensi musi bervariasi, mulai dari 40 Hz hingga 19.5 kHz. Suara lingkungan seperti suara kendaraan, air mengalir, bel pintu, dan masih banyak lagi. Suara-suara tersebut tersebar di seluruh rentang frekuensi yang dapat didengar (Sharma G et al, 2020).

2.2 Musik

Musik adalah seni yang melibatkan pengaturan nada dan suara dalam urutan, kombinasi, dan hubungan temporal untuk menciptakan komposisi yang harmonis dan ekspresif. Musik berfungsi sebagai sarana komunikasi yang universal, mampu menyampaikan emosi dan pesan tanpa batasan bahasa atau budaya (Sujana & Kholid, 2022). Musik dapat dihasilkan melalui berbagai alat musik dan vokal, serta mencakup berbagai genre yang mencerminkan keragaman budaya dan perkembangan teknologi. Dalam era digital saat ini, musik juga telah bertransformasi menjadi produk yang dapat diakses secara luas melalui platform streaming, memungkinkan pendengar untuk menikmati berbagai jenis musik dengan mudah.

2.3 Autotune

Auto-Tune adalah sebuah teknologi yang digunakan untuk memperbaiki ketepatan nada pada vokal penyanyi, sekaligus menambahkan efek suara yang khas. Teknologi ini dirancang untuk mengoreksi nada yang tidak selaras secara otomatis, sehingga menghasilkan performa vokal yang lebih presisi (Nindianto & Panigoro, 2024). Dalam dunia musik, penggunaan Auto-Tune sudah menjadi hal yang umum dan diterima. Bahkan, banyak artis terkenal yang mengandalkan Auto-Tune untuk memperkuat kualitas vokal mereka atau menciptakan efek suara yang unik dan modern. Selain itu, Auto-Tune juga sering digunakan sebagai alat kreatif dalam berbagai genre musik, seperti pop, hip-hop, dan EDM, untuk memberikan karakteristik suara yang ikonik dan inovatif.

2.4 Feature Extraction

Data audio terdiri dari *signal* berdimensi tinggi yang mencakup informasi kompleks. oleh sebab itu diperlukan proses untuk menyederhanakan informasi-informasi tersebut, salah satunya melalui proses *feature extraction*. *Feature extraction* merupakan proses ekstraksi fitur numerik yang merepresentasikan fitur-fitur pada data audio (Tzanetakis, 2002). Fitur-fitur yang umum diekstrak dari data audio adalah fitur spektral. Fitur tersebut meliputi MFCC, spectral centroid, dan chroma.

2.4.1 MFCC

Mel-Frequency Cepstral Coefficients atau MFCC merupakan salah satu teknik feature extraction. Nilai MFCC didapatkan melalui lima proses, yaitu signal framing, penghitungan power spectrum (DFT), penerapan filterbank mel pada power spectrum, penghitungan nilai logaritma filterbank, dan yang terakhir adalah penghitungan DCT (Abdul & Al-Talabani, 2022). MFCC memiliki kelebihan mampu mengekstraksi data dengan tidak menghilangkan atau mengurangi informasi penting yang ada (Ajinurseto & Islamuddin, 2023). Dengan kelebihan tersebut, MFCC biasanya digunakan dalam menganalisis data dari berbagai bidang, seperti bidang akustik & musik, kesehatan, dan industri.

2.4.2 Spectral Centroid

Spectral centroid merupakan penyederhanaan dari nilai distribusi power spectrum (Kato, dkk., 2014). Sumber lain menyebutkan spectral centroid merupakan rata-rata tertimbang frekuensi dengan nilai bobotnya adalah daya pada frekuensi tersebut (Elzaafarany, dkk., 2019). Secara sederhana, spectral centroid dapat dikatakan sebagai frekuensi yang mengandung energi maksimum dari sinyal audio. Metode feature extraction ini memiliki kelebihan antara lain sensitif terhadap frekuensi, mudah digunakan, dan stabil pada kondisi yang berubah-ubah.

2.4.3 Chroma

Chroma bertujuan untuk merepresentasikan fitur harmonik pada sinyal audio dalam waktu yang singkat (Shah, dkk., 2019). Ekstraksi fitur chroma dilakukan dengan cara mengubah sinyal audio menjadi representasi yang lebih sederhana dengan mengelompokkan frekuensi ke dalam 12 kategori sesuai dengan notasi musik. Proses ekstraksi dilakukan dengan mengubah sinyal audio menjadi frekuensi menggunakan transformasi fourier sehingga memungkinkan untuk dilakukan analisis spektrum (Kim & Kim, 2014).

2.5 Klasifikasi dengan Machine Learning

2.5.1 Random Forest

Random Forest adalah algoritma pembelajaran mesin berbasis ensemble yang digunakan untuk klasifikasi dan regresi. Algoritma ini bekerja dengan membangun beberapa pohon keputusan pada subset data yang berbeda dan menggabungkan hasilnya untuk meningkatkan akurasi prediksi. Keunggulan utama dari Random Forest adalah kemampuannya untuk mengurangi overfitting yang sering terjadi pada pohon keputusan tunggal dan meningkatkan akurasi model, terutama ketika menangani data yang besar dan kompleks (Nugraha dkk., 2022).

2.5.2 Support Vector Machine

Support Vector Machine (SVM) adalah algoritma dalam pembelajaran mesin yang termasuk dalam kategori supervised learning, yang digunakan untuk klasifikasi dan regresi. SVM berfungsi untuk menganalisis data dan mengelompokkannya ke dalam dua kategori atau lebih dengan cara mencari hyperplane terbaik yang memisahkan kelas-kelas tersebut dalam ruang fitur. Hyperplane ini adalah batas keputusan yang memisahkan dua kelas data, dan dapat berupa garis dalam dua dimensi atau bidang dalam tiga dimensi (Irfan dkk., 2018).

Namun, SVM memiliki kekurangan yaitu sulit untuk memilih parameter yang optimal sehingga dapat mempengaruhi akurasi (Putri dkk., 2023).

2.5.3 Multi Layer Perceptron

Multi Layer Perceptron (MLP) adalah jenis arsitektur jaringan saraf tiruan dengan banyak lapisan neuron, dimana setidaknya satu lapisan tersembunyi di antara lapisan input dan output (Fatayer & Azara, 2019). MLP adalah jenis jaringan saraf feedforward, yang berarti algoritma ini menghitung jumlah tertimbang dari input dan kemudian menerapkan fungsi aktivasi untuk menghasilkan output. Jaringan ini bekerja dengan mengoptimalkan bobot melalui proses pelatihan, sehingga dapat mengenali pola yang lebih kompleks dalam data. Algoritma ini mampu melakukan klasifikasi pada dataset multiclass dengan baik (Ardiyansyah & Jayanta, 2021).

BAB III METODOLOGI

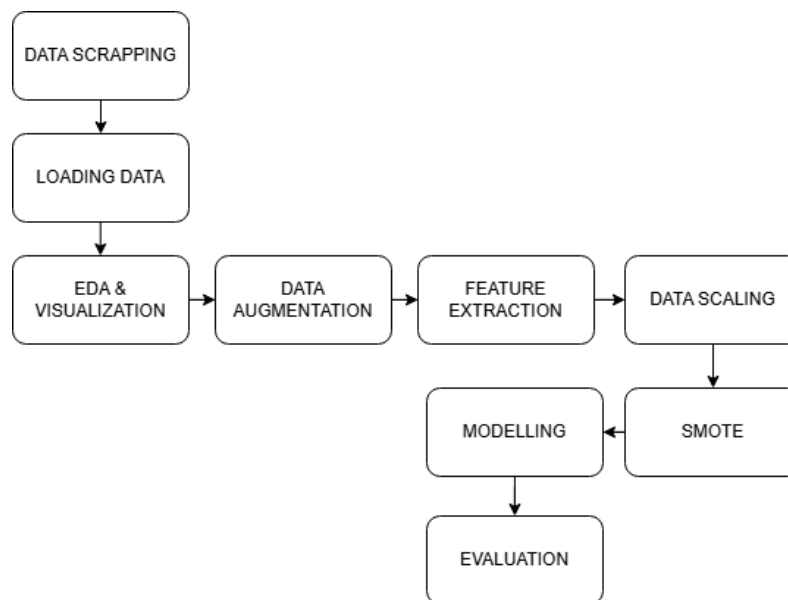
3.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data audio yang berasal dari *website* Freesound. Proses pengumpulan data dilakukan menggunakan *web scraping* menggunakan bahasa pemrograman Python agar mengefisiensi waktu. Dari proses tersebut, didapatkan sebanyak 275 data audio autotune dan 151 data audio no autotune dalam format .wav untuk dilakukan analisis klasifikasi.

Tabel 3.1 Variabel Penelitian

Variabel	Keterangan
<i>Autotune</i>	Data audio dengan suara asli tanpa adanya proses penambahan <i>autotune</i>
<i>No Autotune</i>	Data audio dengan suara yang melalui proses <i>autotune</i>

3.2 Metodologi



Gambar 1. Alur Penelitian

Dalam melakukan analisis klasifikasi audio berdasarkan emosi, terdapat beberapa langkah analisis yang dilakukan.

3.2.1 Pengumpulan data

Hal utama sebelum melakukan analisis klasifikasi audio adalah mengumpulkan data audio yang akan digunakan. Pengumpulan data dapat dilakukan dengan dua cara, yaitu mengambil sendiri data audio yang ada di sekitar atau internet dan cara yang kedua adalah mengambil dataset yang sudah tersedia di internet. Dalam penelitian ini, data yang dikumpulkan adalah dataset yang tersedia di internet dan memberi label untuk masing-masing data audio.

3.2.2 *Eksploratory Data Analysis (EDA)*

Exploratory Data Analysis (EDA) adalah proses analisis data yang dilakukan untuk memahami struktur, pola, dan hubungan dalam data melalui teknik visualisasi dan statistik deskriptif. Tujuannya untuk mengeksplorasi dataset secara menyeluruh sebelum membangun model, sehingga dapat mengidentifikasi anomali, kesalahan, dan hubungan penting antar variabel.

3.2.3 *Preprocessing*

Tahap ini dilakukan untuk mempersiapkan data sebelum mengekstraksi fitur dan modelling. Berdasarkan output EDA, diketahui bahwa frekuensi pada data ini sangat beragam, sehingga dilakukan resampling untuk menyamakan frekuensi data dan memastikan konsistensi dalam analisis. Balancing pada data digunakan untuk menghindari bias pada model yang disebabkan oleh ketidakseimbangan kelas. Selanjutnya, data scaling untuk menormalkan skala fitur dan memastikan semua fitur memiliki rentang nilai yang seragam.

3.2.4 *Ekstraksi fitur audio*

Ekstraksi audio dilakukan untuk mengidentifikasi fitur-fitur audio yang relevan dan menggambarkan karakteristik dari setiap kategori. Dalam penelitian ini, digunakan berbagai fitur relevan untuk audio autotune, seperti MFCC, yang merepresentasikan pola spektral suara, serta fitur lainnya seperti Spectral Centroid, yang menggambarkan kecerahan suara, dan Chroma, yang merepresentasikan tonalitas audio. Selain itu, penelitian ini juga melakukan kombinasi antar fitur dengan tujuan peningkatan akurasi.

3.2.5 *Klasifikasi Model*

Menggunakan algoritma *Support Vector Machine*, *Multilayer Perceptron*, dan *Random Forest* dalam pembuatan model untuk melakukan analisis klasifikasi data audio. Setiap algoritma diuji pada masing-masing fitur untuk mengevaluasi dan menentukan kombinasi fitur yang memberikan hasil terbaik.

3.2.6 *Evaluasi Model*

Terakhir mengukur performa algoritma menggunakan metrik seperti akurasi, precision, recall, dan F1-score. Selain itu, confusion matrix digunakan untuk menganalisis distribusi prediksi model pada setiap kategori. Setiap algoritma dievaluasi berdasarkan fitur seperti MFCC, Spectral Centroid, dan Chroma, serta fitur yang dikombinasikan.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Eksplorasi Data Analisis

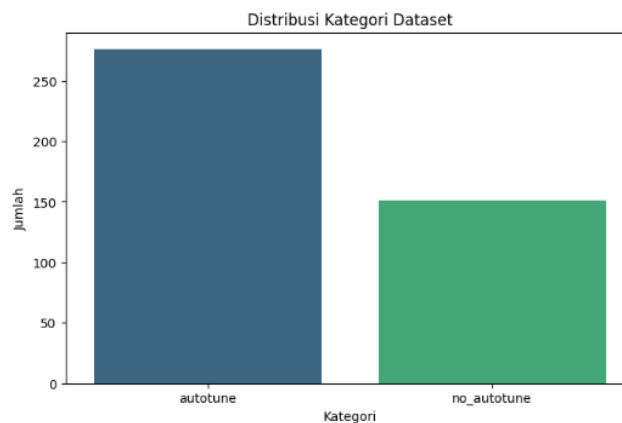
4.1.1 Data Information

Tabel 4.1.1 Data Information

	Autotune	No Autotune
Total File	275	151
Durasi (detik)	5.12	3.89
Sample Rates	96000, 192000, 32000, 48000, 44100, 16000, 24000, 22050	48000, 32000, 96000, 44100

Data Information di atas menjelaskan bahwa terdapat 275 file autotune dan 151 file no autotune. Durasi rata-rata file autotune sebesar 5.12 detik sedangkan no autotune sebesar 3.89. Sample rate untuk kategori no autotune lebih terbatas dibandingkan kategori autotune, yaitu 48000 Hz, 32000 Hz, 96000 Hz dan 44100 Hz.

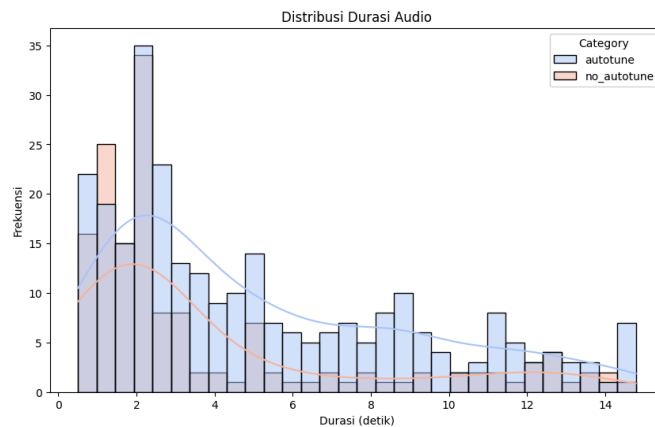
4.1.2 Bar Chart Kategori



Gambar 2. Bar Chart Distribusi Kategori Dataset

Pada bar chart tersebut menggambarkan distribusi jumlah data pada dua kategori, yaitu autotune dan no autotune. Kategori autotune memiliki jumlah data yang lebih banyak, yaitu sebanyak 275 sampel, sedangkan kategori no autotune memiliki jumlah yang jauh lebih kecil, hanya 151 sampel. Hal ini menunjukkan adanya ketidakseimbangan jumlah data antara kedua kategori tersebut.

4.1.4 Distribusi Durasi Audio



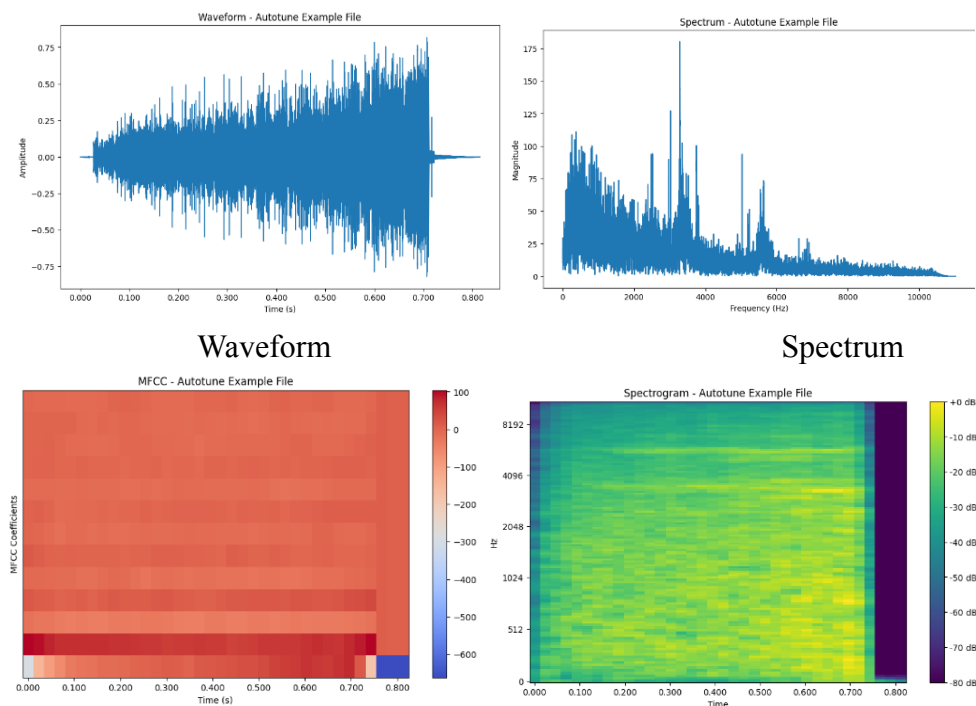
Gambar 3. Distribusi Durasi Audio

Grafik diatas menunjukkan distribusi durasi audio pada dua kategori, yaitu autotune dan no autotune. Sebagian besar sampel pada kedua kategori memiliki durasi pendek, dengan nilai di bawah 5 detik, dengan puncak distribusi pada rentang durasi 0 hingga 2 detik. Frekuensi sampel dengan durasi lebih panjang, khususnya di atas 10 detik, menunjukkan penurunan yang signifikan, terutama pada kategori no autotune. Kategori autotune memiliki jumlah sampel yang lebih tinggi pada seluruh rentang durasi dibandingkan dengan kategori no autotune, yang mengindikasikan ketidakseimbangan jumlah data antar kategori. Distribusi durasi pada kategori no autotune lebih terkonsentrasi pada durasi pendek, sedangkan kategori autotune menunjukkan distribusi yang lebih merata meskipun frekuensinya menurun dengan meningkatnya durasi.

4.1.5 Visualisasi Fitur

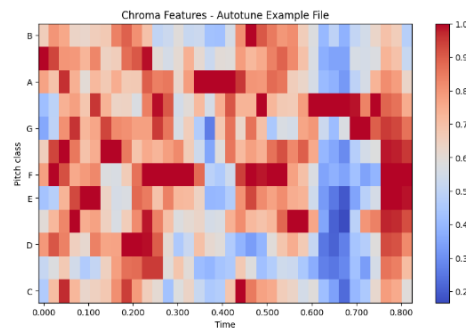
Visualisasi di atas menunjukkan analisis audio untuk kelas "Autotune" menggunakan file dengan indeks ke-0 atau file pertama.

4.1.5.1 Autotune



MFCC

Spectrogram

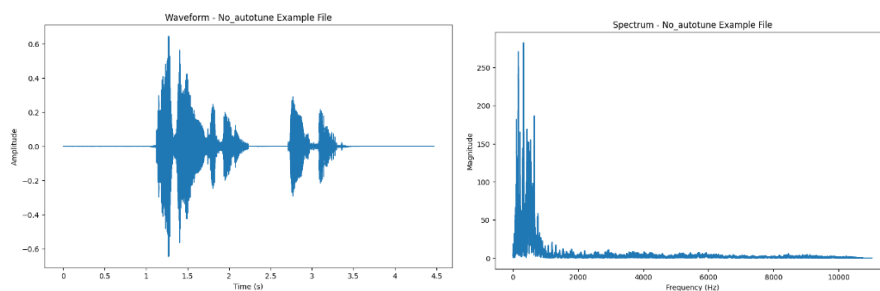


Chroma

Gambar 4. Visualisasi Fitur pada Autotune

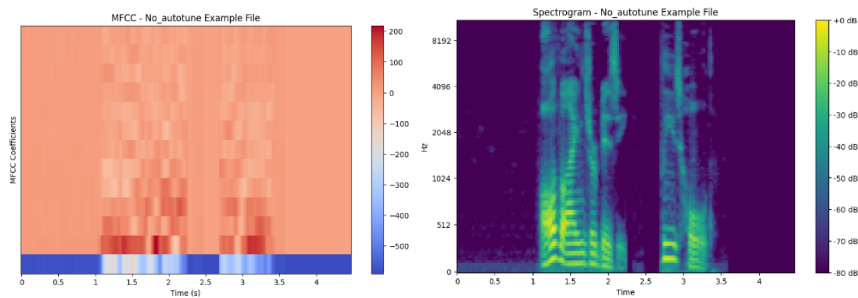
Waveform kelas ‘Autotune’ menunjukkan bahwa suara autotune memiliki fluktuasi energi yang konsisten, mencerminkan efek pengolahan suara yang halus namun terstruktur. Kemudian pada spectrum adanya puncak-puncak tajam di beberapa frekuensi tertentu, menunjukkan bahwa suara autotune memiliki nada dominan atau pitch yang ditonjolkan, artinya adanya manipulasi suara untuk menciptakan frekuensi spesifik yang menonjol. MFCC (Mel-Frequency Cepstral Coefficients) menampilkan pola fitur suara yang cenderung menekankan frekuensi tertentu untuk menciptakan efek tonal khas. Spectrogram autotune memperlihatkan distribusi energi yang terpusat pada rentang frekuensi tertentu, menonjolkan pola waktu yang konsisten, menunjukkan stabilitas dan pengendalian efek autotune pada suara. Sementara itu, chroma menunjukkan intensitas yang merata pada beberapa nada, menunjukkan bahwa efek autotune sering digunakan untuk menciptakan penyesuaian tonal tertentu yang mendukung pola pitch spesifik.

4.1.5.2 No Autotune



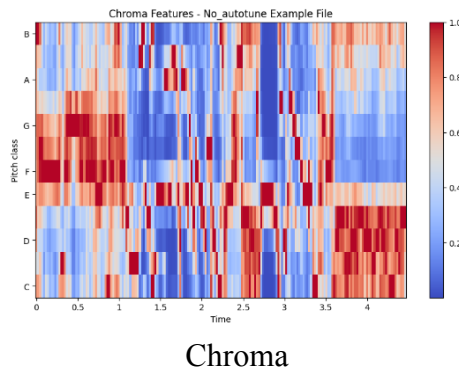
Waveform

Spectrum



MFCC

Spectrogram

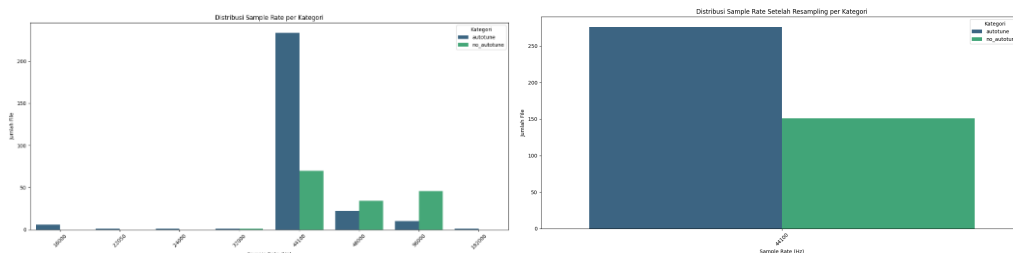


Gambar 5. Visualisasi Fitur pada no autotune

Waveform kelas ‘No Autotune’ menunjukkan fluktuasi amplitudo yang lebih variatif, mencerminkan dinamika suara yang lebih alami dan tidak mengalami pengolahan signifikan. Pola energi ini menunjukkan bahwa suara tidak dimodifikasi untuk menciptakan stabilitas tonal yang diatur. Pada Spectrum, distribusi intensitas frekuensi cenderung lebih merata tanpa adanya puncak dominan yang menonjol, mengindikasikan bahwa suara tidak dimanipulasi untuk memperkuat nada atau pitch tertentu, sehingga karakteristik suara tetap orisinal. MFCC (Mel-Frequency Cepstral Coefficients) memperlihatkan pola fitur suara yang lebih acak dan kompleks, mencerminkan distribusi energi frekuensi yang lebih alami tanpa penekanan pada frekuensi tertentu. Spectrogram menunjukkan distribusi energi frekuensi terhadap waktu yang lebih beragam, tanpa pola waktu yang konsisten seperti pada autotune, sehingga suara memiliki karakteristik yang lebih dinamis dan kurang terstruktur. Chroma menampilkan intensitas pitch yang bervariasi dengan pola yang tidak seragam, menunjukkan bahwa suara tidak diatur untuk menghasilkan harmoni tonal tertentu.

4.2 Preprocessing

4.2.1 Data Augmentation



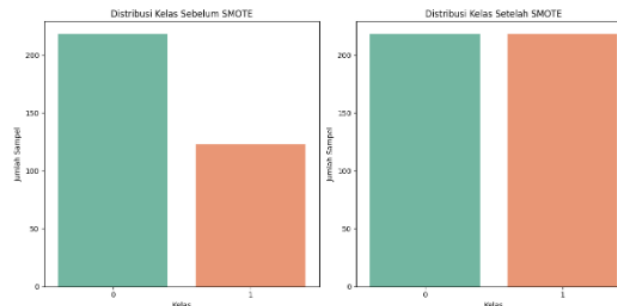
Gambar 6. Data Augmentation

Berdasarkan bar chart di atas, distribusi sample rate pada dataset menunjukkan variasi yang signifikan, dengan sebagian besar file audio di kategori autotune memiliki sample rate 44100 Hz (233 file), namun terdapat juga file dengan sample rate yang berbeda, seperti 96000 Hz (10 file), 48000 Hz (22 file), dan sebagainya, begitu juga pada kategori no_autotune. Sehingga dilakukan resampling untuk menyamakan sample rate dari seluruh data audio menjadi 44100 Hz. Resampling dilakukan untuk mengubah semua file ke sample rate yang konsisten, yaitu 44100 Hz, yang merupakan standar umum dalam pengolahan audio dan memungkinkan mengurangi ketidakkonsistenan yang dapat mempengaruhi performa model.

4.2.2 Normalisasi

Penelitian ini juga melakukan normalisasi terhadap setiap fitur yang digunakan untuk memastikan bahwa semua fitur berada dalam skala yang sama. Normalisasi dilakukan dengan menggunakan fungsi scaling, yang mengubah nilai-nilai fitur agar memiliki rentang yang seragam

4.2.3 SMOTE



Gambar 7. Kelas Sebelum dan Sesudah SMOTE

Berdasarkan visualisasi di atas, terlihat pada bar chart sebelum SMOTE kelas 0 memiliki jumlah sampel yang jauh lebih banyak dibandingkan Kelas 1, menunjukkan ketidakseimbangan data dan dapat menyebabkan model machine learning bias terhadap kelas mayoritas. Setelah SMOTE diterapkan, jumlah sampel pada kedua kelas menjadi seimbang.

4.3 Evaluasi Kinerja Feature Extraction

Pada sub bab ini, hasil analisis mengenai kinerja model berdasarkan evaluasi beberapa teknik ekstraksi yang diuji, yaitu MFCC, spectral centroid, chroma, dan beberapa kombinasi fitur yang berbeda. Evaluasi diukur menggunakan empat metrik utama, yaitu akurasi, precision, recall, dan f1-score.

4.3.1 Random Forest

Tabel di bawah menunjukkan bahwa fitur Chroma memberikan kinerja terbaik, dengan Precision, Recall, dan F1-Score hampir mencapai nilai 1.00, baik pada sinyal audio autotune maupun sinyal audio tanpa autotune. Di sisi lain, MFCC menunjukkan kinerja yang cukup baik, terutama pada sinyal audio autotune, meskipun sedikit menurun pada sinyal audio tanpa autotune. Fitur Spectral Centroid memberikan hasil yang lebih rendah, terutama pada sinyal audio tanpa autotune, yang mengindikasikan bahwa fitur ini kurang efektif dibandingkan ekstraksi fitur yang lain. Kombinasi fitur seperti MFCC + Spectral Centroid dan MFCC + Chroma juga menunjukkan hasil yang lebih baik daripada fitur tunggal, dengan kinerja terbaik tercatat pada MFCC + Spectral Centroid + Chroma dengan autotune, yang menghasilkan nilai Accuracy 0.90. Secara keseluruhan, perbedaan ekstraksi fitur yang digunakan terbukti berpengaruh terhadap hasil akurasi model Random Forest.

Tabel 4.3.1 Random Forest

Feature Extraction	Autotune			No Autotune			Accuracy
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	
MFCC	0.84	0.92	0.88	0.77	0.63	0.69	0.83

Spectral Centroid	0.77	0.73	0.75	0.47	0.52	0.49	0.66
Chroma	1.00	0.99	0.99	0.98	1.00	0.99	0.99
Feature Manipulation	0.92	0.95	0.93	0.88	0.82	0.85	0.91
Feature Inversion	0.91	0.84	0.88	0.72	0.82	0.77	0.84
MFCC + Spectral Centroid	0.89	0.95	0.92	0.87	0.74	0.80	0.88
MFCC + Chroma	0.87	0.90	0.88	0.76	0.70	0.73	0.84
MFCC + Spectral Centroid + Chroma	0.89	0.97	0.93	0.91	0.74	0.82	0.90

4.3.2 Support Vector Machine

Tabel di bawah berisi informasi hasil evaluasi model Support Vector Machine. Feature Manipulation menunjukkan hasil terbaik dengan Accuracy 0.87, terutama pada data sinyal Autotune. Feature Inversion juga memberikan hasil yang baik, dengan Accuracy 0.84. Untuk penggunaan fitur tunggal, MFCC menunjukkan hasil yang baik pada sinyal audio autotune, meskipun turun signifikan pada sinyal audio tanpa autotune, terutama pada Precision dan Recall. Spectral Centroid dan Chroma memiliki kinerja lebih rendah, terutama pada sinyal audio autotune. Kombinasi MFCC + Spectral Centroid + Chroma memberikan hasil yang cukup baik, dengan Accuracy 0.76, meskipun tidak setinggi teknik manipulasi fitur.

Tabel 4.3.2 Support Vector Machine

Feature Extraction	Autotune			No Autotune			Accuracy
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	
MFCC	0.79	0.64	0.71	0.45	0.63	0.52	0.64
Spectral Centroid	0.72	0.61	0.66	0.36	0.48	0.41	0.57
Chroma	0.71	0.68	0.70	0.49	0.53	0.51	0.62
Feature Manipulation	0.89	0.93	0.91	0.84	0.75	0.79	0.87
Feature Inversion	0.88	0.88	0.88	0.75	0.75	0.75	0.84
MFCC + Spectral Centroid	0.82	0.76	0.79	0.55	0.63	0.59	0.72

MFCC + Chroma	0.84	0.69	0.76	0.51	0.70	0.59	0.70
MFCC + Spectral Centroid + Chroma	0.84	0.80	0.82	0.60	0.67	0.63	0.76

4.3.3 Multi-Layer Perceptron

Tabel di bawah ini merupakan hasil mengevaluasi performa model Multi-Layer Perceptron (MLP) dengan berbagai teknik ekstraksi fitur, baik pada sinyal audio autotune maupun sinyal audio tanpa autotune. Hasilnya menunjukkan bahwa Chroma memberikan kinerja terbaik, dengan Accuracy mencapai 0.94, serta Precision, Recall, dan F1-Score yang sangat tinggi. Fitur MFCC juga menunjukkan performa yang baik, dengan Accuracy 0.78, meskipun tidak setinggi Chroma. Sebaliknya, Spectral Centroid menunjukkan hasil yang lebih rendah, dengan Accuracy 0.66. Teknik Feature Manipulation menghasilkan Accuracy 0.83, lebih baik dibandingkan dengan Feature Inversion. Penggabungan fitur, seperti MFCC + Chroma dan MFCC + Spectral Centroid + Chroma, memberikan hasil yang sedikit lebih baik tetapi tidak signifikan dibandingkan dengan penggunaan fitur individu dengan nilai Accuracy 0.81.

Tabel 4.3.3 Multi-Layer Perceptron

Feature Extraction	Autotune			No Autotune			Accuracy
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	
MFCC	0.86	0.81	0.83	0.63	0.70	0.67	0.78
Spectral Centroid	0.81	0.66	0.73	0.47	0.67	0.55	0.66
Chroma	0.97	0.93	0.95	0.89	0.94	0.91	0.94
Feature Manipulation	0.86	0.88	0.87	0.74	0.71	0.73	0.83
Feature Inversion	0.85	0.79	0.82	0.62	0.71	0.67	0.77
MFCC + Spectral Centroid	0.85	0.90	0.88	0.75	0.67	0.71	0.83
MFCC + Chroma	0.85	0.90	0.88	0.75	0.67	0.71	0.83
MFCC + Spectral Centroid + Chroma	0.86	0.86	0.86	0.70	0.70	0.70	0.81

BAB V

KESIMPULAN

5.1 Kesimpulan

Kesimpulan yang dapat diambil berdasarkan analisis dan pembahasan diatas adalah sebagai berikut:

1. Hasil visualisasi fitur menunjukkan bahwa kelas ‘Autotune’ memiliki suara yang terstruktur dengan pola energi stabil, pitch dominan, dan harmoni tonal hasil manipulasi. Sebaliknya, kelas ‘No Autotune’ mencerminkan suara alami dengan fluktuasi energi variatif, distribusi frekuensi merata, dan karakteristik dinamis tanpa manipulasi.
2. Berdasarkan hasil evaluasi yang dilakukan, algoritma Random Forest menunjukkan performa terbaik dalam membedakan audio musik yang menggunakan autotune, dengan memperoleh akurasi tertinggi yaitu 0,90. Dibandingkan dengan algoritma lain, Support Vector Machine hanya mencapai akurasi sebesar 0,76, sementara Multi-Layer Perceptron menghasilkan akurasi 0,81. Hal ini menunjukkan bahwa algoritma Random Forest lebih efektif dalam melakukan klasifikasi audio yang mengandung autotune dibandingkan dengan kedua algoritma lainnya.
3. Berdasarkan hasil evaluasi kinerja model dapat ditarik kesimpulan bahwa pemilihan metode feature extraction yang tepat dapat mempengaruhi kinerja model. Sehingga, penggunaan suatu feature extraction yang baik pada suatu model belum tentu baik jika diterapkan pada model lainnya.

5.2 Saran

Saran untuk penelitian selanjutnya mencakup penerapan teknik augmentasi data pada data audio, seperti pitch shifting dan time-stretching, untuk meningkatkan robustitas model. Penggunaan metode tersebut dapat membantu model menjadi lebih tahan terhadap variasi dalam data yang diolah, sehingga memungkinkan model untuk belajar lebih baik dan mengenali pola yang berbeda dalam data audio. Selain itu, disarankan melakukan uji coba dengan dataset yang lebih besar. Penggunaan dataset yang lebih besar dapat membantu mengidentifikasi sejauh mana model mampu menggeneralisasi performa pada data yang belum pernah dilihat sebelumnya, serta mengurangi potensi overfitting pada data training.

DAFTAR PUSTAKA

- Abdul, Z. K., & Al-Talabani, A. K. (2022). Mel frequency cepstral coefficient and its applications: A review. *IEEE Access*, 10, 122136-122158.
- Abirawa, I. D., Osmond, A. B., & Setianingsih, C. (n.d.). *Klasifikasi emosi pada lirik lagu menggunakan metode Support Vector Machine* [Emotion classification in a song lyrics using Support Vector Machine method]. Program Studi S1 Sistem Komputer, Fakultas Teknik Elektro, Universitas Telkom.
- Ajinurseto, G., & Islamuddin, N. (2023). Penerapan Metode Mel Frequency Cepstral Coefficients pada Sistem Pengenalan Suara Berbasis Desktop. *INFOMATEK: Jurnal Informatika, Manajemen dan Teknologi*, 25(1), 11-20.
- Ardiansyah, D., Jayanta. (2021). Model Klasifikasi Emosi berdasarkan Suara Manusia dengan Metode Multilayer Perceptron. *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, 2(1), 689-702.
- Atahan, Y., Elbir, A., Keskin, A. E., Kiraz, O., Kirval, B., & Aydin, N. (2021, October). Music genre classification using acoustic features and autoencoders. In *2021 Innovations in Intelligent Systems and Applications Conference (ASYU)* (pp. 1-5). IEEE.
- Elzaafarany, K., Aly, M. H., Kumar, G., & Nakhmani, A. (2019). Cerebral artery vasospasm detection using transcranial doppler signal analysis. *Journal of Ultrasound in Medicine*, 38(8), 2191-2202.
- Fatayer, T. S., & Azara, M. N. (2019). IoT secure communication using ANN classification algorithms. In *2019 International Conference on Promising Electronic Technologies (ICPET)* (pp. 142–146).
- Fox, M., Vaidyanathan, G., & Breese, J. L. (2024). The impact of artificial intelligence on musicians. *Issues in Information Systems*, 25(3).
- Gohari, M., Bestagini, P., Benini, S., & Adami, N. (2024). Spectrogram-Based Detection of Auto-Tuned Vocals in Music Recordings. *arXiv preprint arXiv:2403.05380*.
- Hartayu, R., Santoso., Anam, C., & Aziz, D.A. (2022). Simulasi Ekstraksi Fitur Suara menggunakan Mel-Frequency Cepstrum Coefficient. *Jurnal Sains dan Informatika*, 8(1), 80-87. DOI: 10.34128/jsi.v8i1.357
- Kato, K., Nagao, T., Yamanaka, T., Kawai, K., & Sakakibara, K. I. (2014). Effect of room acoustics on timbral brightness of clarinet tones: Experimental investigation with two binaural room impulse responses. *Acoustical Science and Technology*, 35(6), 300-308.
- Kim, H. G., & Kim, J. Y. (2014). Robust audio fingerprinting method using prominent peak pair based on modulated complex lapped transform. *ETRI Journal*, 36(6), 999-1007.
- Laksana, E. A., & Sulianta, F. (2017). Analisis dan studi komparatif algoritma klasifikasi genre musik. *Semnasteknomedia Online*, 5(1), 2-1.
- McFee, B., Raffel, C., & others. (2024). Feature extraction. *librosa*. <https://librosa.org/doc/latest/feature.html>
- Mustaqim, Y., Utami, E., & Raharjo, S. (2019). KLASIFIKASI AUDIO MENGGUNAKAN WAVELET TRANSFORM DAN NEURAL NETWORK. *Informasi Interaktif*, 4(2), 122-130.

- Nindianto, A. (2024). *The usage of Auto-Tune and pitch correction technology in the music industry and advertising* (Skripsi, Universitas Indonesia). Fakultas Ilmu Sosial dan Ilmu Politik Universitas Indonesia.
- Nugraha, A. F., Aziza, R. F. F., & Pristyanto, Y. (2022). Penerapan metode Stacking dan Random Forest untuk meningkatkan kinerja klasifikasi pada proses deteksi web phishing. *Jurnal Infomedia: Teknik Informatika, Multimedia & Jaringan*, 7(1), 39.
- Putri, T. B., Saidah, S., Hidayat, B., Qothrunnada, F., & Darwindra, D. (2023). Deteksi emosi berdasarkan sinyal suara manusia menggunakan discrete wavelet transform (DWT) dengan klasifikasi support vector machine (SVM). *Jurnal Ilmu Komputer dan Informatika (JIKI)*, 3(1), 1–10. <https://doi.org/10.54082/jiki.45>
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan machine learning dalam berbagai bidang. *Jurnal Khatulistiwa Informatika*, 5(1), 490845.
- Shah, A., Kattel, M., Nepal, A., & Shrestha, D. (2019). Chroma feature extraction. Chroma Feature Extraction using Fourier Transform.
- Sujana, A., & Kholid, D. M. (n.d.). Proses produksi musik pada lagu *Lara* oleh band Dialog Senja. *SWARA: Jurnal Antologi Pendidikan Musik*. Retrieved November 26, 2024, from <https://ejournal.upi.edu/index.php/antomusik/index>
- Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing*, 10(5), 293-302.
- Viznyak, N. (2024). Harmonizing Minds: Understanding Grade 10 Students' Perspectives on AI-Generated Music. Available at SSRN 4830842.

Lampiran

Link

Dataset:

<https://drive.google.com/drive/folders/1POPfZS0UT0eG81eBgj1WcGP1mXbXfoLH>

Link pengerjaan code modelling:

https://colab.research.google.com/drive/185OlSFzcBgCdZBKSw_kgeh0TkhXa0Onk?usp=sharing

Link

pengerjaan

EDA:

<https://colab.research.google.com/drive/1W4ETwOajd-FEA4onbdyGkXQyHLPp8KiO>