

IF2123 ALJABAR LINEAR DAN GEOMETRI

**Implementasi Music Information Retrieval (MIR) dan Principal  
Component Analysis (PCA) dalam Sistem Pencarian Audio  
Berdasarkan Input Gambar dan Humming untuk Meningkatkan  
Akurasi Retrieval**

**Laporan Tugas Besar 2**

Disusun untuk memenuhi tugas mata kuliah Aljabar Linear dan Geometri pada  
Semester 1 (satu)

Tahun Akademik 2024/2025



**Kelompok Boboiboy Kuasa Tiga**

**Oleh:**

**Frederiko Eldad      13523147**

**Benedictus Nelson    13523150**

**Theo Kurniady        13523154**

**PROGRAM STUDI TEKNIK INFORMATIKA  
SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA  
INSTITUT TEKNOLOGI BANDUNG  
BANDUNG  
2024**

# DAFTAR ISI

<b>DAFTAR ISI.....</b>	<b>2</b>
<b>BAB I.....</b>	<b>3</b>
1.1 Latar Belakang Masalah.....	3
1.2 Spesifikasi Program.....	4
<b>BAB II.....</b>	<b>5</b>
2.1 Image Retrieval.....	5
2.2 Sistem Temu Balik Suara (MIR).....	8
<b>BAB III.....</b>	<b>11</b>
3.1 Arsitektur Website (frontend).....	11
3.2 Arsitektur Program Information Retrieval (backend).....	11
<b>BAB IV.....</b>	<b>13</b>
4.1 Studi Kasus #1 : File Midi.....	13
4.2 Studi Kasus #2 : File WAV.....	13
4.3 Studi Kasus #3 Bagian lagu.....	14
4.4 Studi Kasus #4 : File gambar.....	14
<b>BAB V.....</b>	<b>15</b>
5.1 Kesimpulan.....	15
5.2 Komentar.....	16
5.3 Refleksi.....	16
<b>LAMPIRAN.....</b>	<b>17</b>

# BAB I

## DESKRIPSI MASALAH

### 1.1 Latar Belakang Masalah

Gambar adalah salah satu bentuk data visual yang paling sering digunakan manusia untuk memahami dunia di sekitarnya. Dalam kehidupan sehari-hari, kita menggunakan gambar untuk mengidentifikasi objek, memahami pola, dan menyimpan kenangan. Dalam dunia teknologi, gambar juga menjadi bagian penting yang dapat dianalisis untuk berbagai keperluan, seperti pencarian gambar serupa, pengenalan wajah, dan deteksi objek. Proses pencarian gambar serupa, atau yang dikenal sebagai *Image Retrieval*, menggunakan metode matematik untuk mengubah data visual menjadi bentuk representasi yang dapat diproses oleh komputer. Teknik seperti *grayscale image*, *Euclidean Distance Calculation*, dan *Principal Component Analysis (PCA)* sering digunakan untuk menyederhanakan data gambar tanpa kehilangan informasi penting, sehingga memungkinkan pencocokan gambar dilakukan dengan lebih cepat dan efisien.

Sama halnya dengan gambar, suara juga merupakan elemen tak terpisahkan dari kehidupan manusia. Suara digunakan untuk berkomunikasi, mengekspresikan emosi, dan menikmati seni, seperti musik. Dalam teknologi modern, suara dapat dianalisis dan dikenali untuk berbagai kebutuhan, seperti identifikasi lagu, pengenalan suara, hingga analisis percakapan. Salah satu teknologi berbasis suara yang populer adalah aplikasi pengenalan lagu, seperti Shazam, yang mampu mengidentifikasi lagu hanya dari cuplikan pendek suara. Kemampuan ini didasarkan pada teknik *Music Information Retrieval (MIR)*, yang memungkinkan sistem untuk mengekstrak ciri khas suara, seperti frekuensi, tempo, dan ritme, untuk kemudian dibandingkan dengan data yang ada di basis data menggunakan metode seperti *histogram dan cosine similarity*.

Dalam tugas besar ini, konsep *image retrieval* dan *audio retrieval* akan digabungkan untuk menciptakan sistem yang mampu mengenali gambar dan suara dengan akurasi tinggi. Teknik seperti PCA akan digunakan untuk mengurangi kompleksitas data, sementara metode penghitungan kesamaan akan membantu dalam proses pencocokan. Dengan memahami dan mengimplementasikan teknologi ini, sistem pencarian yang lebih efisien dapat dikembangkan, memungkinkan penerapan luas mulai dari hiburan hingga kebutuhan industri modern.

## **1.2 Spesifikasi Program**

Program ini dikembangkan dengan menggunakan arsitektur website yang dirancang untuk memberikan pengalaman pengguna yang sederhana namun efektif. Antarmuka yang sederhana memudahkan pengguna untuk mengakses dan menikmati setiap fitur yang tersedia, baik dalam konteks Image Retrieval maupun Music Information Retrieval. Setiap fitur diimplementasikan sesuai dengan kebutuhan spesifik dari kedua aplikasi tersebut, memastikan program dapat memberikan hasil yang akurat dan relevan. Selain itu, program ini juga dilengkapi dengan kemampuan untuk menganalisis pola data dan fungsi menggunakan pendekatan matematis yang beragam, memberikan pengguna solusi yang komprehensif dalam memproses dan mengevaluasi informasi.

### **1.2.1 Bahasa Pemrograman**

Bahasa pemrograman yang digunakan dalam pengembangan tugas besar ini adalah typescript dan python. Pemilihan typescript didasarkan pada penggunaan framework Next.JS dalam pembuatan website bagian frontend dan python didasarkan pada berbagai library yang mempermudah perhitungan. Frontend menggunakan framework Next.JS dan tailwind CSS untuk membentuk tampilan website. Back end menggunakan library flask pada python untuk menyambungkan API dari program perhitungan dengan website.

### **1.2.2 Library**

Library yang kami gunakan pada implementasi PCA dan MIR adalah library PIL, numpy, mido, librosa, dan os. Library numpy digunakan dalam perhitungan seperti rata-rata dan histogram. Library mido digunakan untuk ekstraksi file jenis .midi atau .mid. Library librosa digunakan untuk ekstraksi file jenis .wav. Library os digunakan untuk mengakses directory. Library PIL digunakan dalam memproses gambar.

# BAB II

## TEORI SINGKAT

### 2.1 Image Retrieval

*Image Retrieval* adalah teknik pengolahan data visual yang bertujuan untuk mencari gambar tertentu dari kumpulan gambar yang tersimpan dalam database. Awalnya, pencarian gambar dilakukan dengan memberikan metadata pada setiap gambar di database. Pengguna akan memasukkan query berupa kata-kata untuk menemukan gambar tersebut. Pendekatan ini dikenal sebagai *Text-Based Image Retrieval* (TBIR). Namun, TBIR memiliki beberapa kelemahan, seperti kurangnya relevansi gambar yang ditemukan karena metadata yang tidak akurat dan membutuhkan ruang penyimpanan khusus untuk metadata.

Untuk mengatasi kekurangan TBIR, dikembangkan metode lain, yaitu *Content-Based Image Retrieval* (CBIR). Teknik ini dianggap lebih unggul karena tidak lagi bergantung pada kata-kata, melainkan pada fitur-fitur yang terdapat dalam gambar itu sendiri. Pengguna dapat memberikan query berupa gambar yang akan di-*retrieve*. CBIR terbukti lebih relevan dibandingkan TBIR karena pencarian dilakukan berdasarkan fitur seperti warna, bentuk, dan tekstur gambar. Salah satu metode CBIR yang populer adalah metode PCA (*Principal Component Analysis*) menggunakan konsep SVD (*Singular Value Decomposition*).

Langkah-langkah untuk melakukan *image retrieval* menggunakan PCA adalah sebagai berikut:

#### 2.1.1 Image Processing and Loading

Untuk mengurangi kompleksitas gambar dan memfokuskan analisis pada bagian terang dan gelap gambar, gambar akan diubah menjadi *grayscale*. Ini berarti setiap gambar direpresentasikan dalam intensitas piksel saja tanpa informasi warna.

$$I(x, y) = 0.2989 \cdot R(x, y) + 0.5870 \cdot G(x, y) + 0.1140 \cdot B(x, y)$$

Selanjutnya, semua gambar disesuaikan ukurannya menjadi seragam untuk memastikan konsistensi dan meningkatkan akurasi perhitungan.

Terakhir, gambar *grayscale* yang awalnya berupa matriks berdimensi  $M \times N$  diubah menjadi vektor satu dimensi dengan panjang  $M \cdot N$ , sehingga data dapat diproses lebih mudah dalam format vektor.

$$I = [I_1, I_2, \dots, I_{M \cdot N}]$$

### 2.1.2 Data Centering (Standardization)

Pada langkah ini, data distandarisasi di sekitar nilai 0. Caranya adalah dengan menghitung rata-rata setiap gambar untuk suatu piksel dengan formula:

$$\mu_j = \frac{1}{N} \sum_{i=1}^N x_{ij}$$

di mana:

- $x_{ij}$  adalah nilai piksel ke-j pada gambar ke-i,
- $N$  adalah jumlah total gambar dalam dataset.

Setelah itu, setiap nilai piksel dikurangi dengan rata-rata piksel yang telah dihitung untuk melakukan standarisasi, menggunakan formula:

$$x_{ij'} = x_{ij} - \mu_j$$

### 2.1.3 PCA Computation Using Singular Value Decomposition (SVD)

Setelah gambar-gambar distandarisasi, maka akan dilakukan perhitungan SVD (*Singular Value Decomposition*) pada gambar-gambar tersebut. Perhitungan SVD dimulai dengan membuat matriks kovarians dari data-data yang sudah distandarisasi. Pembuatan matriks kovarians dilakukan dengan menggunakan formula:

$$C = \frac{1}{N} X'^T X'$$

di mana:

- $X'$  adalah matriks yang sudah distandarisasi.

Setelah itu, akan dilakukan dekomposisi nilai singular untuk mendapatkan komponen utama menggunakan formula:

$$C = U \Sigma U^T$$

di mana:

- $U$  adalah matriks *eigenvector* (komponen utama),
- $\Sigma$  adalah matriks *eigenvalue* (menunjukkan varian data di sepanjang komponen utama).

Lalu, akan diambil  $n$  jumlah komponen utama teratas dari hasil SVD dan dilakukan proyeksi gambar ke komponen utama. Proyeksi data dilakukan dengan memilih  $k$ -komponen utama teratas ( $k \ll M \cdot N$ ) dan menggunakan formula:

$$Z = X'U_k$$

di mana:

- $U_k$  adalah matriks *eigenvector* dengan  $n$ -dimensi.

#### 2.1.4 Similarity Computation

Pencarian kesamaan gambar query dengan gambar di database dilakukan dengan mencari jarak euclidean di antaranya. Untuk mencari jarak euclidean, gambar query direpresentasikan terlebih dahulu dalam ruang komponen utama dengan proyeksi yang sama menggunakan formula:

$$q = (q' - \mu)U_k$$

di mana:

- $q$  adalah vektor proyeksi dari gambar query ke ruang komponen utama (PCA),
- $q'$  adalah gambar query dalam format vektor (setelah grayscale, resize, dan flattening),
- $\mu$  adalah rata-rata piksel dari dataset (per piksel),
- $U_k$  adalah matriks *eigenvector* dengan  $k$  dimensi utama dari PCA.

Kemudian, jarak Euclidean antara gambar query dengan semua gambar dalam dataset dihitung dengan formula:

$$d(q, z_i) = \sqrt{\sum_{j=1}^k (q_j - z_{ij})^2}$$

di mana:

- $d(q, z_i)$  adalah jarak antara gambar query  $q$  dan gambar ke- $i$  dalam ruang komponen utama,
- $z_i$  adalah vektor proyeksi dari gambar ke- $i$  dalam dataset ke ruang komponen utama,
- $q_j$  adalah elemen ke- $j$  dari vektor proyeksi query  $q$ ,
- $z_{ij}$  adalah elemen ke- $j$  dari vektor proyeksi gambar ke- $i$ , yaitu  $z_i$ ,
- $k$  adalah jumlah dimensi ruang komponen utama yang dipilih.

Lalu, jarak diurutkan dari yang terkecil.

#### 2.1.5 Retrieval and Output

Gambar-gambar yang mirip dengan query akhirnya dikumpulkan dengan cara melakukan limitasi jumlah atau dengan memberikan batas jarak euclidean.

## 2.2 Sistem Temu Balik Suara (MIR)

Music Information Retrieval (MIR) adalah cabang ilmu komputer yang berfokus pada teknik-teknik untuk mengakses dan menganalisis informasi yang terkandung dalam data suara, khususnya data musik. Dalam konteks ini, sistem temu balik suara bertujuan untuk memungkinkan pengguna menemukan informasi atau lagu berdasarkan data audio yang ada. Proses ini melibatkan ekstraksi fitur audio, seperti frekuensi, durasi, dan nada, untuk memodelkan dan membandingkan musik secara lebih efisien.

Teknologi ini banyak digunakan dalam aplikasi seperti pencarian musik, rekomendasi lagu, serta pengenalan musik dari input pengguna, termasuk fitur seperti query by humming atau song recognition. Dalam MIR, pendekatan yang digunakan mencakup analisis spektrum suara dan perhitungan kesamaan antara data audio, yang memungkinkan hasil pencarian yang lebih relevan dan akurat.

### 2.2.1 Metode Ekstraksi fitur

Ekstraksi fitur merupakan langkah penting dalam proses Music Information Retrieval (MIR). Pada tahap ini, berbagai jenis fitur audio diekstraksi untuk membantu dalam menganalisis dan membandingkan lagu atau suara. Beberapa metode yang sering digunakan dalam ekstraksi fitur meliputi:

- **Absolute Tone Based (ATB)**

Fitur ini mengukur nada absolut dalam sebuah lagu dengan menilai nada yang dihasilkan berdasarkan frekuensi dasar. Fitur ini penting untuk menangkap elemen-elemen utama dari sebuah lagu, seperti notasi atau akor yang dimainkan

- **Relative Tone Based (RTB)**

Metode ini lebih fokus pada perbandingan antara nada yang satu dengan yang lainnya. RTB digunakan untuk mengetahui pola relatif dalam musik, misalnya interval antara dua nada atau perubahan nada dalam sebuah melodi.

- **First Tone Based (FTB)**

FTB digunakan untuk menilai nada pertama dalam sebuah segmen audio yang bisa membantu dalam mengenali karakteristik dasar sebuah lagu, khususnya dalam query by humming.



Selain itu, fitur-fitur tersebut diolah menggunakan teknik histogram untuk menggambarkan distribusi frekuensi fitur-fitur ini dalam data audio.

### 2.2.2 Normalisasi

Setelah ekstraksi fitur dilakukan, langkah selanjutnya adalah normalisasi. Normalisasi bertujuan untuk memastikan bahwa nilai-nilai dalam histogram berada pada skala probabilitas, yang memungkinkan perbandingan yang adil antar fitur. Dalam konteks ini, histogram yang dihasilkan dari proses ekstraksi fitur dinormalisasi agar totalnya menjadi 1, yang merupakan sifat dari distribusi probabilitas.

Berikut adalah formula umum untuk normalisasi histogram:

$$H_{norm}[d] = \frac{H[d]}{\sum_{d=0}^{127} H[d]}$$

Dimana:

- $H$  adalah histogram
- $d$  adalah bin pada histogram

Normalisasi ini memastikan bahwa setiap nilai dalam histogram terukur dalam skala yang konsisten.

### 2.2.3 Teknik Windowing

Salah satu teknik yang digunakan dalam pemrosesan audio adalah windowing, yang membagi sinyal audio menjadi segmen-segmen kecil. Pada umumnya, dalam metode ini, setiap window berukuran antara 20 hingga 40 pitch dengan sliding window antara 4 hingga 8 pitch. Sliding window yang lebih kecil memberikan akurasi yang lebih tinggi, karena lebih banyak informasi yang dapat diekstraksi dari sinyal audio, meskipun membutuhkan lebih banyak waktu komputasi. Dengan teknik ini, setiap segmen dapat dianalisis secara terpisah untuk memudahkan identifikasi pola dan fitur dalam audio.

#### 2.2.4 Perhitungan Kemiripan dengan Cosine Similarity

Setelah proses ekstraksi dan normalisasi fitur, langkah selanjutnya adalah perhitungan kemiripan antara audio atau query dengan data yang ada. Dalam tugas ini, Cosine Similarity digunakan sebagai metode untuk mengukur seberapa mirip dua vektor fitur.

Cosine Similarity mengukur sudut antara dua vektor dalam ruang berdimensi tinggi. Semakin kecil sudut antara kedua vektor (semakin mendekati nilai 1), semakin mirip kedua vektor tersebut. Formula dari Cosine Similarity adalah sebagai berikut:

$$\text{Cosine Similarity}(A, B) = \frac{A \cdot B}{||A|| ||B||}$$

Dimana:

- $A$  dan  $B$  adalah dua vektor fitur yang akan dibandingkan.
- $A \cdot B$  adalah perkalian titik antara vektor  $A$  dan  $B$ .
- $||A||$  dan  $||B||$  adalah panjang (norma) dari masing-masing vektor.

Penerapan Cosine Similarity dalam MIR memungkinkan pencarian musik yang lebih akurat dengan membandingkan vektor fitur hasil ekstraksi dan normalisasi, serta menentukan kemiripan antara query dan data dalam database.

# BAB III

## ARSITEKTUR SISTEM WEB DAN BACKEND

### 3.1 Arsitektur Website (frontend)

#### 3.1.1. Frameworks

Dalam pembuatan frontend dari website ini, tech stack / frameworks yang digunakan adalah Next.js sebagai *core framework*, @headlessui/react untuk *styling* komponen, Tailwind CSS untuk *styling* CSS, midi-player-js dan Soundfont Player untuk memutar audio dengan ekstensi .mid / .midi, dan React Icons untuk ikon yang digunakan dalam tampilan frontend website.

#### 3.1.2. Components

Komponen yang dibuat untuk digunakan dalam frontend pada website ini adalah *Card Component* yaitu audiocard, audiosearch, imagesearch, dan *Side Navigation Bar* yaitu sidebar. Masing-masing komponen dibuat menggunakan TypeScript Execute (TSX).

Komponen audiocard merupakan komponen yang dapat digunakan kembali, dan digunakan dalam audiosearch dan imagesearch. Komponen ini akan menampilkan sebuah *container* yang berisi nama file, gambar, skor kemiripan (tidak ditampilkan), dan *handler onPlay* untuk file audio.

Komponen audiosearch, dan imagesearch merupakan komponen yang akan digunakan pada masing-masing laman pencaharian. Dengan audiosearch memanggil API untuk metode MIR, dan imagesearch memanggil API untuk metode PCA.

Komponen sidebar, merupakan komponen yang digunakan di seluruh laman yang ada sebagai navigasi untuk berpindah laman dan mengunggah file.

### 3.2 Arsitektur Program Information Retrieval (backend)

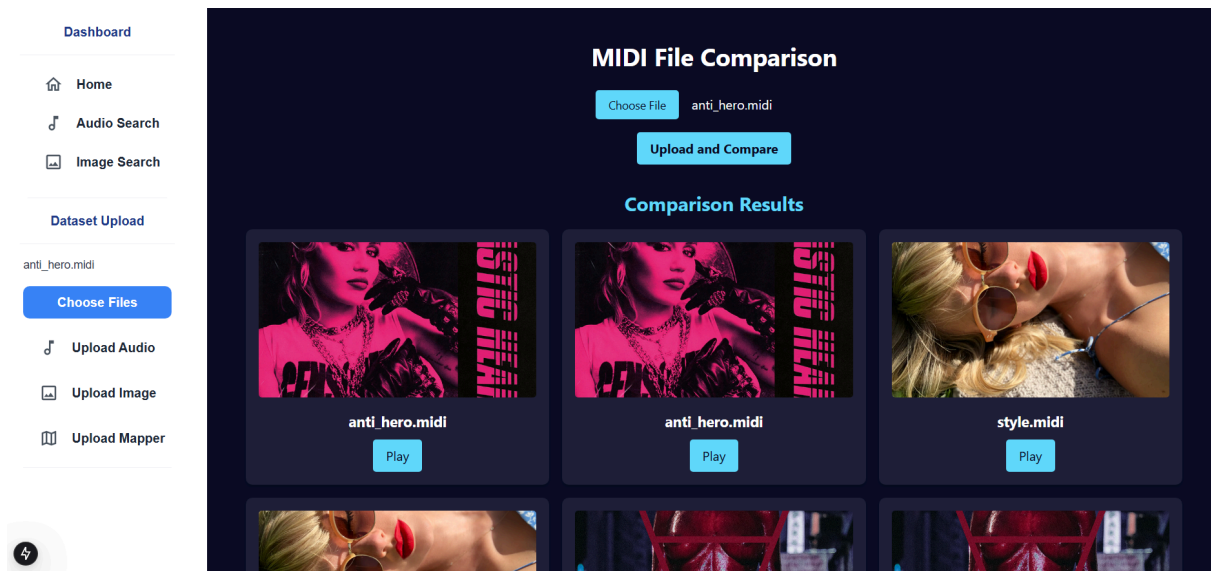
kiwkiw



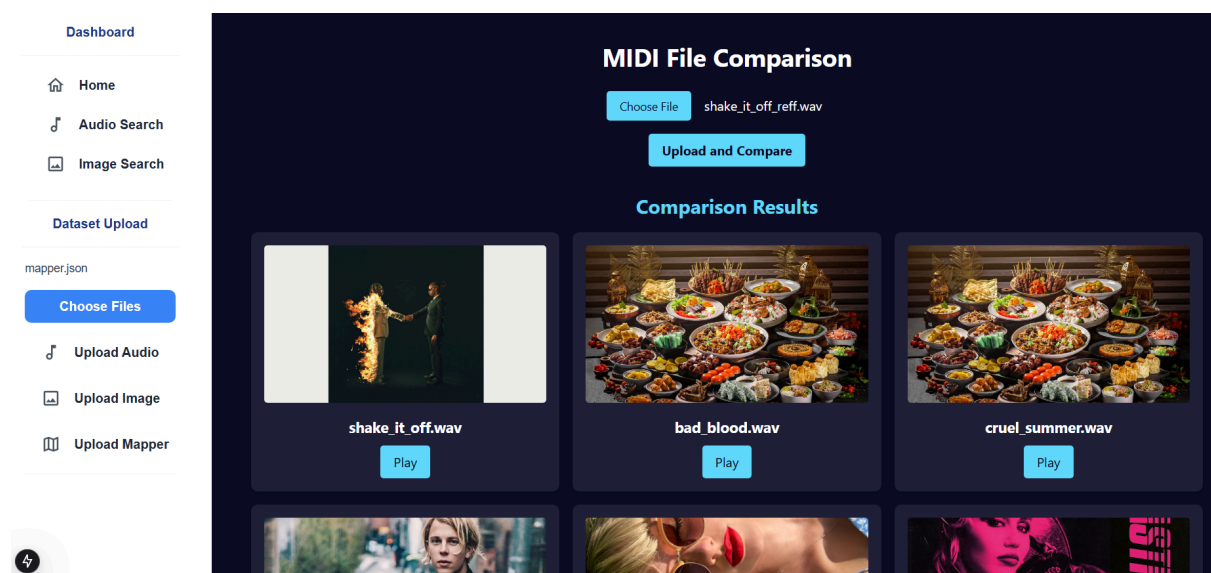
# BAB IV

## EKSPERIMEN

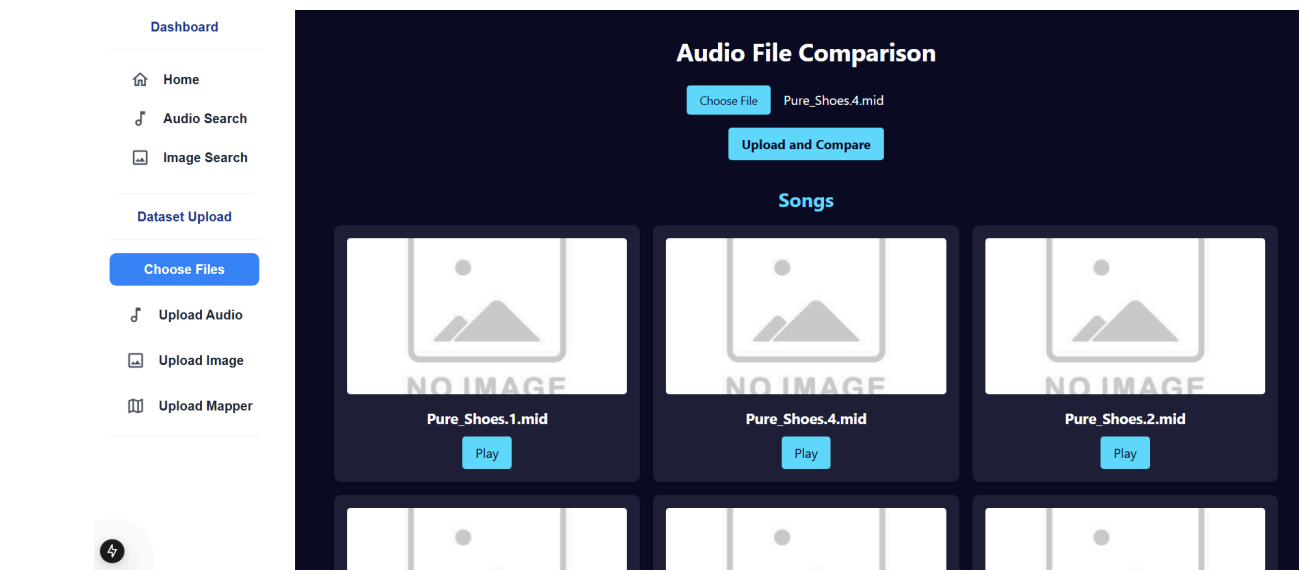
### 4.1 Studi Kasus #1 : File Midi



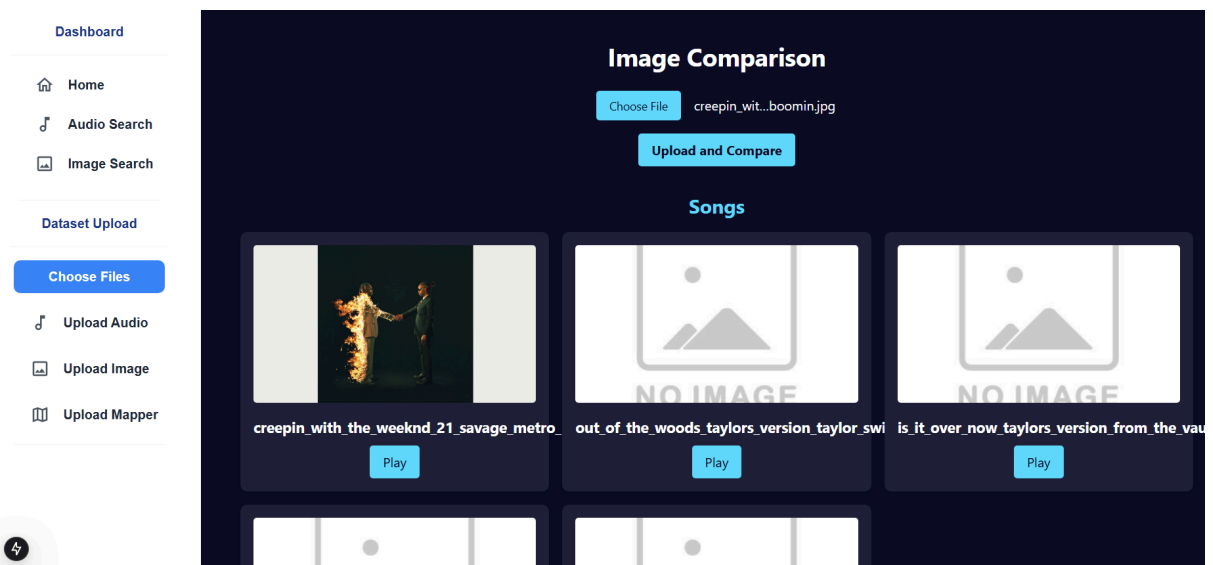
### 4.2 Studi Kasus #2 : File WAV



### 4.3 Studi Kasus #3 Bagian lagu



### 4.4 Studi Kasus #4 : File gambar



# BAB V

## KESIMPULAN

### 5.1 Kesimpulan

Dalam tugas besar ini, kami berhasil mengembangkan program retrieval konten yang mampu digunakan untuk pencarian image dan audio dalam dataset. Program ini dapat melakukan beberapa metode seperti :

- **Image Retrieval** menggunakan *Principal Component Analysis* (PCA) serta perhitungan jarak dengan *Euclidean Distance*
- **Audio Retrieval** dengan beberapa metode ekstraksi :
  - *Absolute Tone Based* (ATB)
  - *Relative Tone Based* (RTB)
  - *First Tone Based* (FTB)Fitur tersebut dihitung dalam histogram, kemudian dibandingkan menggunakan *Cosine Similarity*.
- **Mapper Retrieval** menggunakan *file* berbasis JSON yang menghubungkan file audio dengan image
- **Upload Dataset**: Mendukung pengunggahan dataset audio maupun gambar secara dinamis untuk memperbarui sistem.
- **Integrasi API**: Program menyediakan API modular dengan endpoint terpisah untuk setiap fitur, termasuk pencarian gambar, pencarian audio, dan pemetaan dataset.

Program yang dibangun dapat menerima input secara manual dari file gambar atau audio query (format .jpg, .wav, atau .midi), kemudian memberikan hasil yang akurat sesuai metode retrieval yang dipilih. Program juga mampu menyelesaikan studi kasus nyata, seperti pencarian gambar album berdasarkan gambar input atau pencocokan audio menggunakan metode query by humming.

Selain itu, program ini:

- Diimplementasikan menggunakan arsitektur backend Flask dan frontend React untuk mendukung penggunaan melalui antarmuka web yang interaktif.
- Mendukung optimasi kinerja dengan penggunaan PCA untuk gambar dan segmen pitch histogram untuk audio.
- Dikonversikan menjadi sistem yang modular, memungkinkan pengguna untuk memperluas fitur atau memperbarui dataset dengan mudah.

## 5.2 Komentar

Selama proses pengembangan program ini, kami cukup stres

## 5.3 Refleksi

Proyek ini memberikan kami kesempatan untuk

- **Implementasi materi** : Dengan proyek ini, kami dapat mengimplementasikan secara langsung materi yang telah diajarkan pada kelas Aljabar Linier dan Geometri. Materi yang dimaksud mencakup seluruh materi vektor, dengan tambahan PCA.
- **Web Development** : Dengan proyek ini, kami dapat mempelajari web based development, mulai dari struktur, setup, hingga penggabungan frontend dan backend.

## 5.4 Saran

Saran kami untuk tugas besar Algeo kedepannya adalah untuk lebih mengarah ke perhitungan dengan materi Algeo, bukan web development..



# LAMPIRAN

## HOMEPAGE

