

LAPORAN TUGAS BESAR 2

SISTEM TEMU BALIK SUARA DAN IMAGE RETRIEVAL

DENGAN PCA BERBASIS WEBSITE

Diajukan untuk memenuhi Tugas Besar 2 IF-2123 Aljabar Linier dan Geometri

Dosen pengajar: Dr. Ir. Rinaldi Munir, M.T.



Disusun oleh:

Kelompok Slebew Hanif

Orvin Andika (13523017)

Salman Hanif (13523056)

Aryo Wisanggeni (13523100)

SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA
INSTITUT TEKNOLOGI BANDUNG
BANDUNG 2024

DAFTAR ISI

DAFTAR ISI.....	1
BAB I.....	1
DESKRIPSI MASALAH.....	1
1. Pemrosesan Suara dan Gambar.....	1
2. Information Retrieval.....	2
BAB II.....	3
TEORI SINGKAT.....	3
2.1 Sistem Temu Balik Suara.....	3
2.2 Ekstraksi Fitur Berdasarkan Humming.....	3
2.3 Image Retrieval dengan PCA.....	5
BAB III.....	6
ARSITEKTUR WEBSITE.....	6
3.1 Frontend.....	6
3.2 Backend.....	6
BAB IV.....	7
EKSPERIMENT.....	7
BAB V.....	13
KESIMPULAN.....	13
1. Kesimpulan.....	13
2. Komentar.....	13
3. Saran.....	13
LAMPIRAN.....	14
Referensi.....	14
Tautan.....	14

BAB I

DESKRIPSI MASALAH

1. Pemrosesan Suara dan Gambar

Suara selalu menjadi hal yang paling penting dalam kehidupan manusia. Manusia berbicara mengeluarkan suara dan mendengarkan suatu suara untuk diresap ke otak dan mencari informasi dari suara tersebut. Suara juga bisa dijadikan orang-orang di dunia ini sebuah media untuk membuat karya seni. Contohnya adalah alat mendeteksi lagu. Manusia bisa mendeteksi suara dengan menggunakan indera pendengar dan memberikan kesimpulan akan apa jenis suara tersebut melalui respon dari otak. Sama seperti manusia, teknologi juga bisa mendeteksi suara dan memberikan jawaban mereka melalui algoritma-algoritma yang beragam bahkan bisa melebihi kapabilitas manusia. Dengan menggunakan algoritma apapun, konsep dari pendeksi dan interpretasi suara itu bisa juga disebut dengan sistem temu balik suara atau bisa disebut juga dengan *audio retrieval system*. Banyak aplikasi yang menggunakan konsep sistem temu balik contohnya adalah Shazam.



Gambar 1. Shazam sebagai aplikasi *audio retrieval system*

Selain suara, manusia juga memiliki penglihatan sebagai salah satu inderanya dan bisa melihat warna dan gambar yang bermacam-macam. Teknologi komputasi juga memiliki kapabilitas yang sama dan bisa melihat gambar sama seperti kita, tetapi teknologi seperti ini juga bisa merepresentasikan gambar tersebut sebagai beragam-ragam angka yang bisa disebut juga fitur. Tahun ke tahun, *image processing* selalu menjadi fokus utama dari tugas besar 2 Algeo. Algoritma yang digunakan adalah Eigenvalue, Cosine Similarity, Euclidean Distance, dll.

Anda sudah melewati Tugas Besar 1 yaitu tentang matriks dan implementasi terhadap berbagai hal. Matriks adalah salah satu komponen yang penting dalam aplikasi aljabar vektor. Di dalam Tugas Besar 2 ini, anda diminta untuk membuat semacam aplikasi Shazam yaitu sebuah aplikasi yang meminta input lagu dan aplikasi tersebut mendekripsi apa nama dari lagu tersebut dan beberapa detail lainnya. Pada tugas besar ini, anda akan menggunakan aljabar vektor untuk mencari perbandingan antar satu audio dengan audio yang lain. Anda akan menggunakan konsep yang bernama *Music Information Retrieval* atau MIR untuk mencari dan mengidentifikasi suara berdasarkan fitur-fitur yang dimilikinya. Tidak hanya itu, anda juga

akan menggunakan konsep Principal Component Analysis (PCA) untuk mencari kumpulan audio melalui deteksi wajah berbagai orang (anggap saja mereka sebagai seorang penyanyi).

2. Information Retrieval

Information Retrieval adalah konsep meminta informasi dari sebuah data dengan memasukkan data tertentu. Pada tugas besar ini, anda akan berkutik dengan 2 jenis Information Retrieval. *Image Retrieval* dan *Music Information Retrieval*. *Image Retrieval* adalah konsep untuk memasukkan sebuah input gambar dan berharap mendapatkan gambar yang ada di data sesuai dengan informasi dan perhitungan yang diinginkan. Sedangkan *Music Information Retrieval* (MIR) adalah konsep untuk memasukkan sebuah input audio dan berharap mendapatkan audio yang ada di data sesuai dengan informasi dan perhitungan yang diinginkan. Pada tugas besar kali ini, kalian akan mengimplementasikan *Image Retrieval* dengan menggunakan Principal Component Analysis dan *Music Information Retrieval* dengan menggunakan humming.

BAB II

TEORI SINGKAT

2.1 Sistem Temu Balik Suara

Temu balik informasi adalah pencarian materi tidak terstruktur yang memenuhi kebutuhan informasi dari sejumlah koleksi. Maka dari itu, sistem temu balik suara dapat diartikan sebagai sistem yang menyangkut teknik dan mekanisme untuk merepresentasikan *query* musik, mencari dokumen dalam koleksi musik yang mirip, dan mengembalikan hasil pencarian yang paling mirip dengan *query* pengguna. Metode pencarian pada sistem temu balik suara secara garis besar dibagi menjadi dua yaitu pencarian berdasarkan isi dan pencarian berdasarkan metadata.

Pencarian berdasarkan metadata dapat dikatakan sebagai metode yang lebih sederhana. Secara singkat, metadata adalah sebuah data tentang data. Metadata pada musik dapat berisi tentang judul lagu, judul album, genre, penyanyi, pengarang, tahun rilis, dan lain-lain. Informasi ini disimpan dalam berkas musik agar bisa diakses dimanapun berkas ini disimpan. Sementara itu, pencarian berdasarkan isi merupakan pencarian yang lebih rumit karena menggunakan isi dari musik itu sendiri seperti tempo, ritme, harmoni dan sebagainya.

Pencarian berdasarkan isi ini bisa memanfaatkan *Musical Instrument Digital Interface (MIDI)*. MIDI didefinisikan sebagai standar yang diterima secara universal untuk pertukaran informasi musik secara digital. MIDI menyediakan format digital yang memungkinkan instrumen musik berinteraksi dengan komputer. Format digital ini membuat musik dapat dianalisis pada tingkat yang lebih mendalam seperti identifikasi notasi, struktur melodi, dan pola harmoni.

Penerapan sistem temu balik suara memiliki manfaat luas dalam berbagai bidang, mulai dari layanan streaming musik, pengarsipan musik digital, hingga alat bantu untuk komposer dan produser musik. Dengan semakin berkembangnya teknologi, seperti pembelajaran mesin dan pemrosesan sinyal digital, sistem ini mampu memberikan pengalaman pencarian musik yang semakin presisi dan personal.

2.2 Ekstraksi Fitur Berdasarkan Humming

Ekstraksi fitur berdasarkan humming adalah teknik temu balik suara yang memungkinkan pengguna mencari sebuah lagu hanya dengan menyendungkan melodi dari lagu tersebut. Sistem ini dirancang untuk mencocokkan pola melodi yang dihasilkan dari senandung dengan koleksi lagu dalam basis data berdasarkan analisis isi dari koleksi tersebut. Teknik ini merupakan solusi yang efektif untuk mencari sebuah lagu tanpa mengetahui judul atau metadata lainnya.

Pada query by humming akan dilakukan beberapa langkah utama. Pertama, dilakukan pemrosesan audio, di mana suara yang diinput direkam atau diterima dan dipersiapkan untuk tahap berikutnya. Selanjutnya, data audio tersebut melalui tahap ekstraksi fitur vektor, yaitu proses konversi data audio menjadi representasi numerik berupa vektor fitur yang dapat dianalisis. Terakhir, pada tahap pencarian similaritas vektor, vektor fitur yang dihasilkan dibandingkan dengan dataset untuk menemukan hasil yang paling sesuai atau diatas nilai kemiripan/similaritas yang telah ditentukan.

Pemrosesan audio dalam sistem query by humming menggunakan file MIDI dengan fokus pada track melodi utama, umumnya di Channel 1. Setiap file MIDI diproses menggunakan metode windowing yang membagi melodi menjadi segmen 20-40 beat dengan sliding window 4-8 beat. Teknik ini memungkinkan pencocokan fleksibel dari berbagai bagian lagu yang mungkin diingat pengguna. Proses windowing disertai normalisasi tempo dan pitch untuk mengurangi variasi humming. Setiap note event dikonversi menjadi representasi numerik yang mempertimbangkan durasi dan urutan nada, memungkinkan sistem membandingkan potongan melodi dengan database.

Ekstraksi fitur dilakukan dalam dua tahap. Pertama adalah distribusi *tone*. Distribusi *tone* dilakukan dalam tiga *viewpoints* yaitu *Absolute Tone Based (ATB)*, *Relative Tone Based (RTB)*, dan *First Tone Based (FTB)*. Fitur *Absolute Tone Based (ATB)* menghitung frekuensi kemunculan setiap nada berdasarkan skala MIDI (0-127). Histogram yang dihasilkan memberikan gambaran distribusi absolut nada dalam data. Hal ini penting untuk menangkap karakteristik statis melodi dalam sinyal audio. Langkah pertama adalah membuat histogram dengan 128 bin, sesuai dengan rentang nada MIDI dari 0 hingga 127. Kemudian, hitung frekuensi kemunculan masing-masing nada MIDI dalam data. Setelah itu, normalisasi histogram untuk mendapatkan distribusi yang terstandarisasi. Fitur *Relative Tone Based (RTB)* menganalisis perubahan antara nada yang berurutan, menghasilkan histogram dengan nilai dari -127 hingga +127. RTB berguna untuk memahami pola interval melodi, yang lebih relevan dalam mencocokkan humming dengan dataset yang tidak bergantung pada pitch absolut. Dimulai dengan membangun histogram yang memiliki 255 bin dengan rentang nilai dari -127 hingga +127. Selanjutnya, hitung selisih antara nada-nada yang berurutan dalam data. Terakhir, lakukan normalisasi pada histogram yang telah dibuat. Fitur First Tone Based (FTB) fokus pada perbedaan antara setiap nada dengan nada pertama, menciptakan histogram yang mencerminkan hubungan relatif terhadap titik referensi awal. Pendekatan ini membantu menangkap struktur relatif nada yang lebih stabil terhadap variasi pitch pengguna. Histogram dibuat dengan 255 bin, juga mencakup rentang nilai dari -127 hingga +127. Kemudian, hitung selisih antara setiap nada dalam data dengan nada pertama. Histogram yang dihasilkan kemudian dinormalisasi untuk menghasilkan distribusi yang seimbang. Normalisasi memastikan bahwa semua nilai dalam histogram berada dalam skala probabilitas. Berikut adalah formula umum dari normalisasi yang digunakan:

$$H_{norm} = \frac{H[d]}{\sum_d H[d]}$$

Dimana H adalah Histogram dan d adalah bin dari histogram tersebut.

Penghitungan similaritas dilakukan dengan menghitung cosinus dari dua vektor histogram. Semakin dekat hasil perhitungan cosinus dengan 1, maka kedua suara tersebut semakin mirip. Formula *cosine similarity* adalah sebagai berikut:

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

2.3 Image Retrieval dengan PCA

Image retrieval adalah proses pencarian dan pengambilan gambar dari suatu koleksi berdasarkan fitur tertentu. Salah satu pendekatan yang umum digunakan untuk meningkatkan efisiensi dan akurasi dalam sistem ini adalah Principal Component Analysis (PCA). Principal Component Analysis (PCA) adalah teknik statistik yang digunakan untuk mereduksi dimensi data dengan tetap mempertahankan sebanyak mungkin informasi yang ada. PCA mengubah data berdimensi tinggi menjadi beberapa dimensi yang lebih kecil, disebut principal components, tanpa kehilangan esensi atau pola utama dalam data tersebut. Hasil data yang didapatkan dari PCA ini akan berupa eigenvector dan proyeksi data.

Langkah yang dibutuhkan untuk melakukan pencarian gambar dengan PCA diawali dengan mengubah gambar menjadi grayscale untuk mengurangi kompleksitas gambar dan membuat fokus menjadi bagian terang dan gelap gambar. Selain itu, seluruh gambar pada dataset juga harus memiliki ukuran yang sama. Setelah itu, gambar akan diubah menjadi satu dimensi dan distandarisasi. Matriks kovarian dibentuk dari data yang sudah distandarisasi tersebut. Kemudian, pada matriks kovarian tersebut dilakukan dekomposisi SVD. Lalu diambil n komponen utama dari hasil SVD dan proyeksikan gambar ke komponen utama.

BAB III

ARSITEKTUR WEBSITE

3.1 Frontend

Pada bagian frontend, website ini menyediakan beberapa fitur utama untuk mendukung pengalaman pengguna. Fitur Audio Query Input mencakup form untuk mengunggah file audio yang akan digunakan sebagai query dalam pencocokan suara. Selain itu, terdapat Main Page yang akan menampilkan kumpulan file audio yang telah diunggah ke dalam sistem. Agar navigasi lebih mudah, terutama jika dataset berisi banyak file, fitur pagination digunakan untuk membatasi jumlah file yang ditampilkan per halaman.

Sebagai fitur tambahan, frontend juga mendukung Image Retrieval dengan PCA, di mana file audio yang ada dapat dipetakan ke gambar album terkait. Pengguna bisa melakukan query dengan gambar album sebuah lagu dan website ini akan menampilkan lagu-lagu yang memiliki cover album sama seperti query.

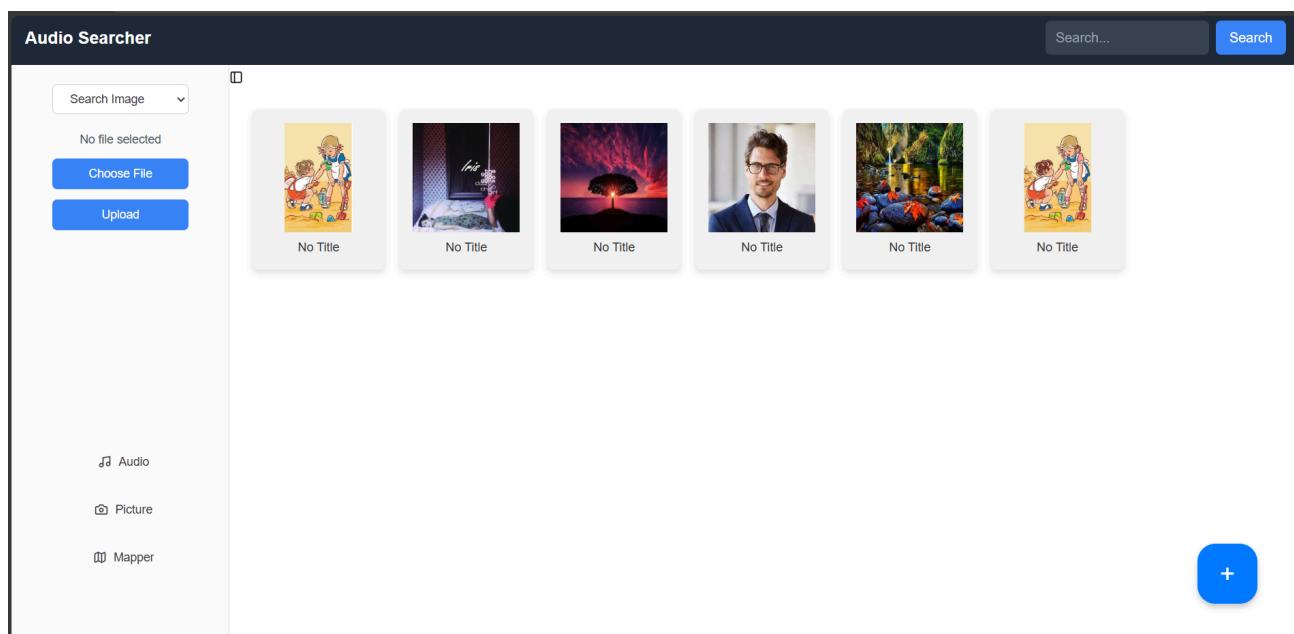
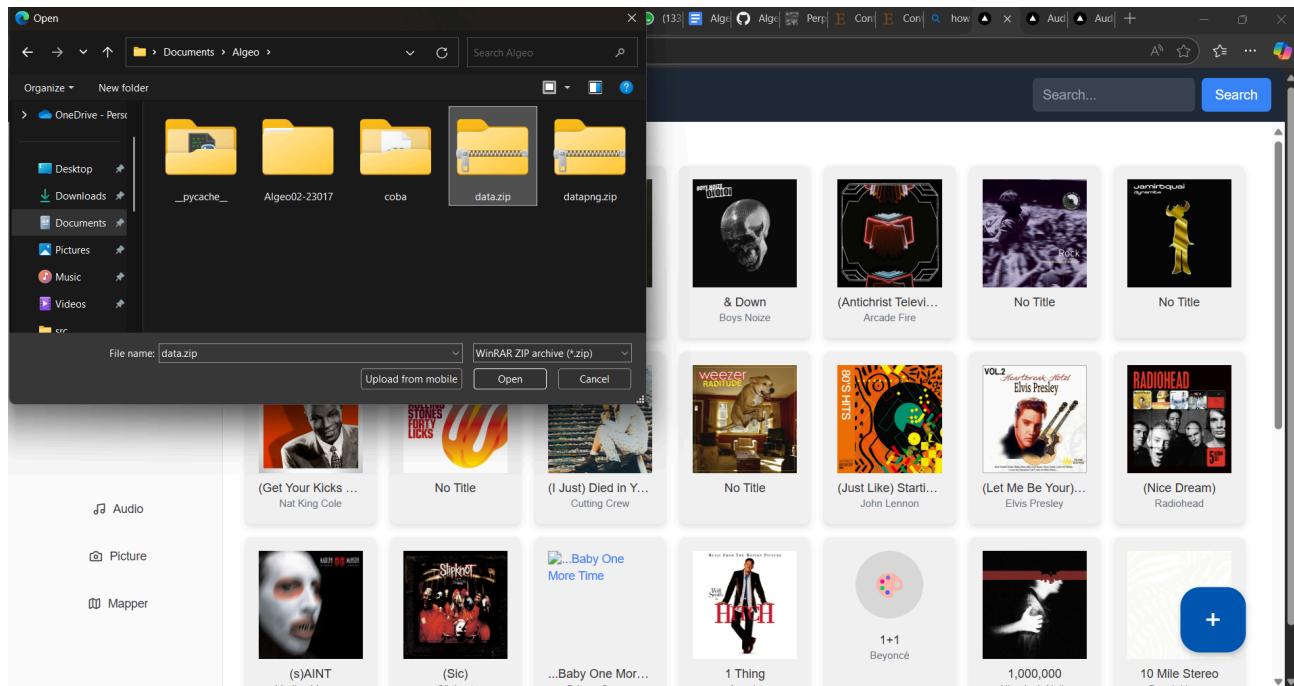
3.2 Backend

Pada bagian backend, website ini menyediakan beberapa fitur utama untuk mendukung pengalaman pengguna. Bagian File Upload Handler di backend dirancang untuk mendukung unggahan berbagai jenis file, termasuk multiple file audio serta file kompresi seperti ZIP dan RAR. Setelah file diunggah, sistem akan secara otomatis mengekstrak file kompresi tersebut dan menyimpan dataset audio yang dihasilkan untuk digunakan dalam proses lebih lanjut. Backend juga memiliki modul Audio Processing & Query Matching yang bertugas memproses file audio baik dari query maupun dataset, kemudian mencocokannya menggunakan algoritma tertentu untuk menemukan suara yang sesuai. Website ini juga memiliki Image Processing & Query Matching yang akan memproses gambar dari dataset dan gambar dan mencocokannya. Selain itu, terdapat modul tambahan Mapper yang digunakan untuk memetakan file audio ke gambar album yang relevan. Proses ini memungkinkan pengguna melakukan query dengan audio dan image untuk mencari sebuah lagu.

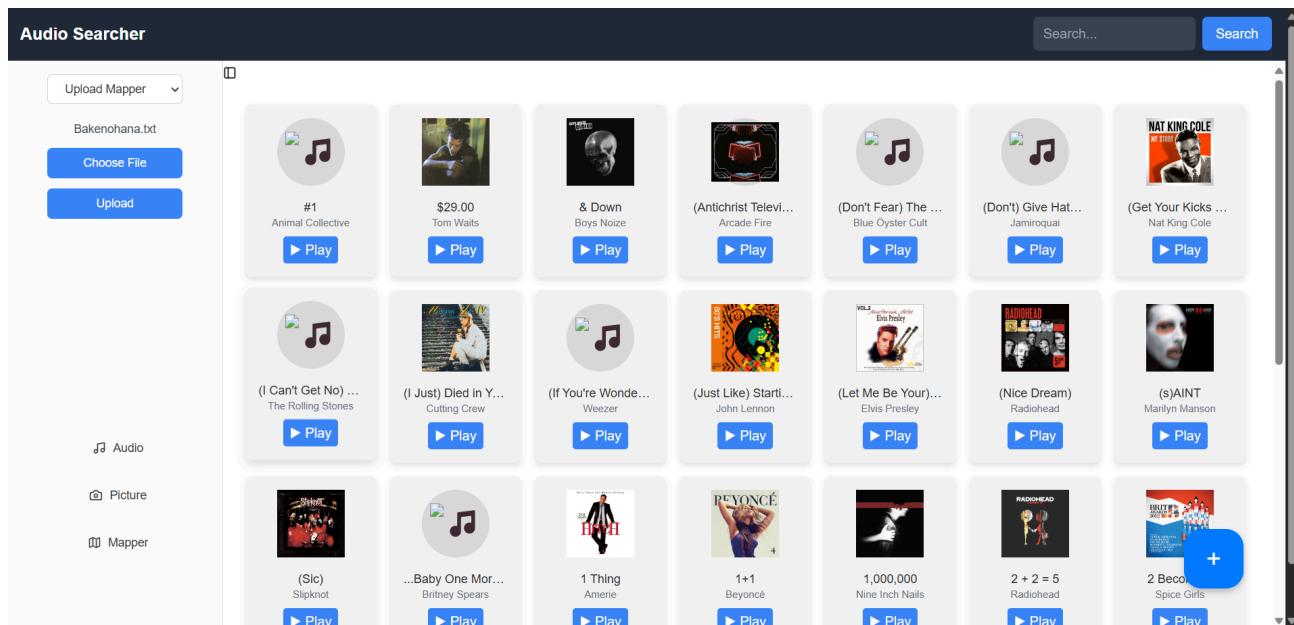
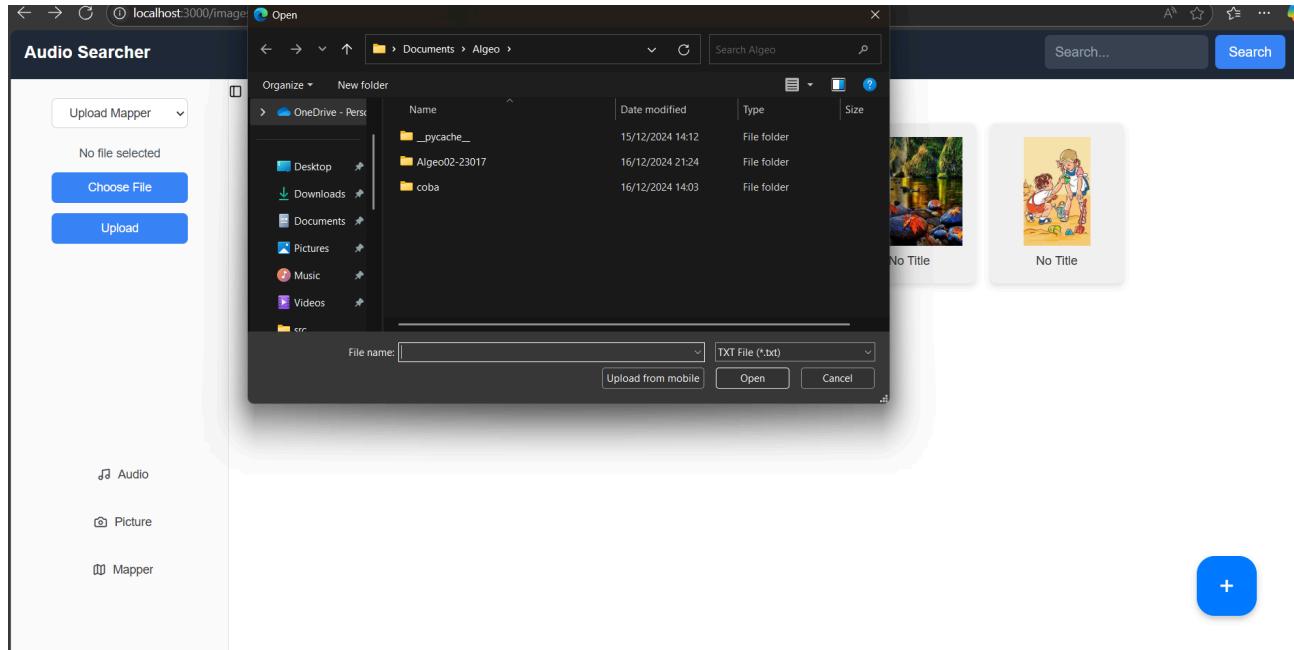
BAB IV

EKSPERIMEN

1. Upload Dataset



2. Upload Mapper



3. Audio Page

Audio Searcher

Search... Search

No file selected

Choose File

Upload

Search Image

Audio Picture Mapper

The screenshot shows a search interface for audio files. On the left, there's a sidebar with options to choose or upload a file, and buttons for 'Search Image', 'Audio', 'Picture', and 'Mapper'. The main area displays a grid of 24 album covers, each with a 'Play' button below it. The albums include: #1 by Animal Collective, \$29.00 by Tom Waits, & Down Boys Noise, (Antichrist Televi...) by Arcade Fire, (Don't Fear) The ... by Blue Oyster Cult, (Don't) Give Hat... by Jamiroquai, (Get Your Kicks ... by Nat King Cole, (I Can't Get No ...) by The Rolling Stones, (I Just) Died in Y... by Cutting Crew, (If You're Wonde... by Weezer, (Just Like) Startl... by John Lennon, (Let Me Be Your)... by Elvis Presley, (Nice Dream) by Radiohead, (s)AINT by Marilyn Manson, (Sic) by Slipknot, ...Baby One Mor... by Britney Spears, 1 Thing by Amerie, 1+1 by Beyoncé, 1,000,000 by Nine Inch Nails, 2 + 2 = 5 by Radiohead, and 2 Become 1 by Spice Girls.

Audio Searcher

Search... Search

No file selected

Choose File

Upload

Search Image

Audio Picture Mapper

This screenshot shows the same interface as above, but with a modal window overlaid on the second row, third column album cover. The modal displays a larger version of the album cover for '\$29.00' by Tom Waits, the title '\$29.00', the artist 'Tom Waits', and a red 'Close' button at the bottom right. The rest of the interface and grid remain visible in the background.

4. Image Page

The screenshot shows a web-based application titled "Audio Searcher" at the URL localhost:3000/images. The interface includes a search bar at the top right with placeholder text "Search...". On the left, there's a sidebar with navigation links: "Search Image" (selected), "Choose File" (button), "Upload" (button), "Audio" (link), "Picture" (link), and "Mapper" (link). The main area displays a grid of album covers. Each item in the grid contains the album cover, the title, and the artist. A blue "+" button is located in the bottom right corner of the grid.

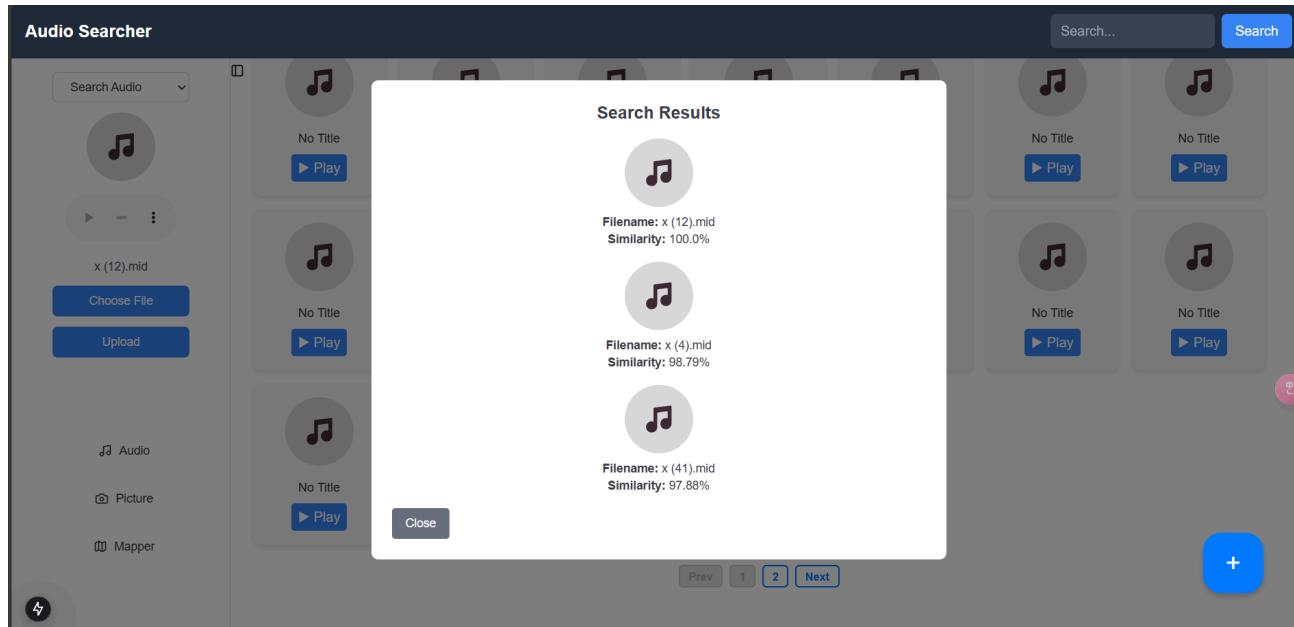
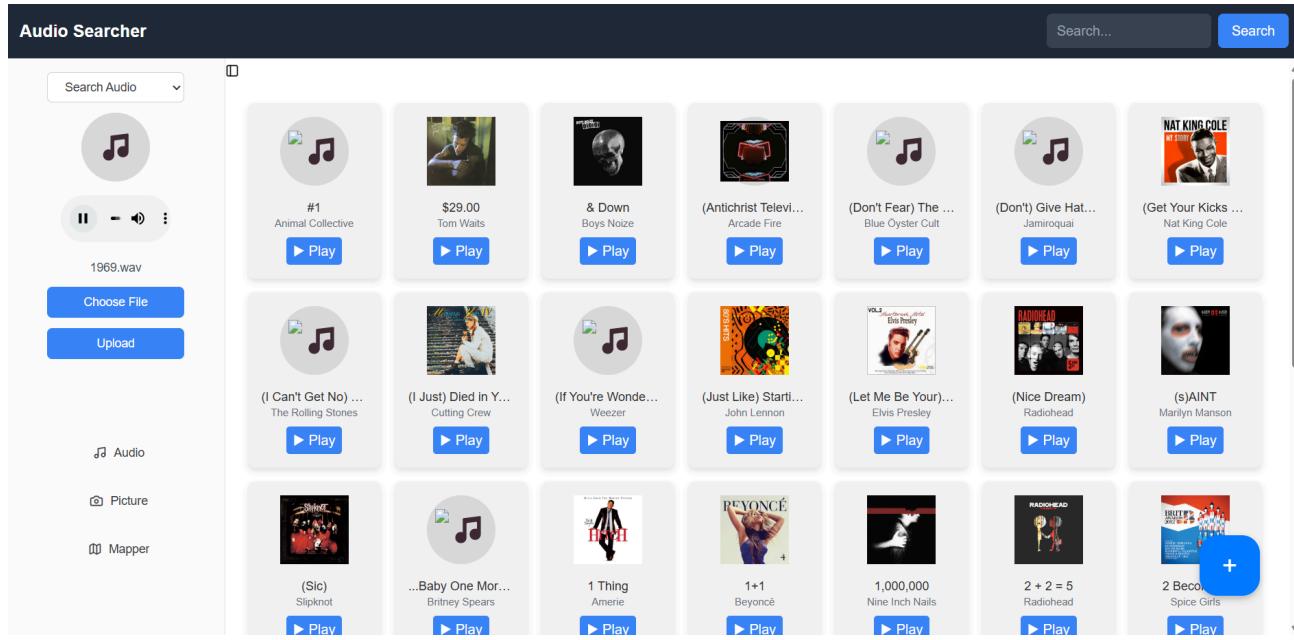
#1	\$29.00	& Down	(Antichrist Televi...	No Title	No Title	NAT KING COLE
Animal Collective	Tom Waits	Boys Noize	Arcade Fire			(Get Your Kicks ... Nat King Cole)
ROLLING STONES	(I Just) Died In Y...	weezer	SUNSOO	VOL.2	RADIOHEAD	(s)AINT
Subtly Licks	Cutting Crew	ELVIS PRESLEY	(Just Like) Startl...	Heartbreak Hotel Elvis Presley	5	Marilyn Manson
No Title	No Title	No Title	John Lennon	(Let Me Be Your)...	(Nice Dream)	
Slipknot	1 Thing	Beyoncé	1,000,000	10 Mile Stereo	10 Years Today	11th Hour
(Sic)	Amerie		Nine Inch Nails	Reach House	Bullet for My Valentine	Lamb of God

5. Mapper Page

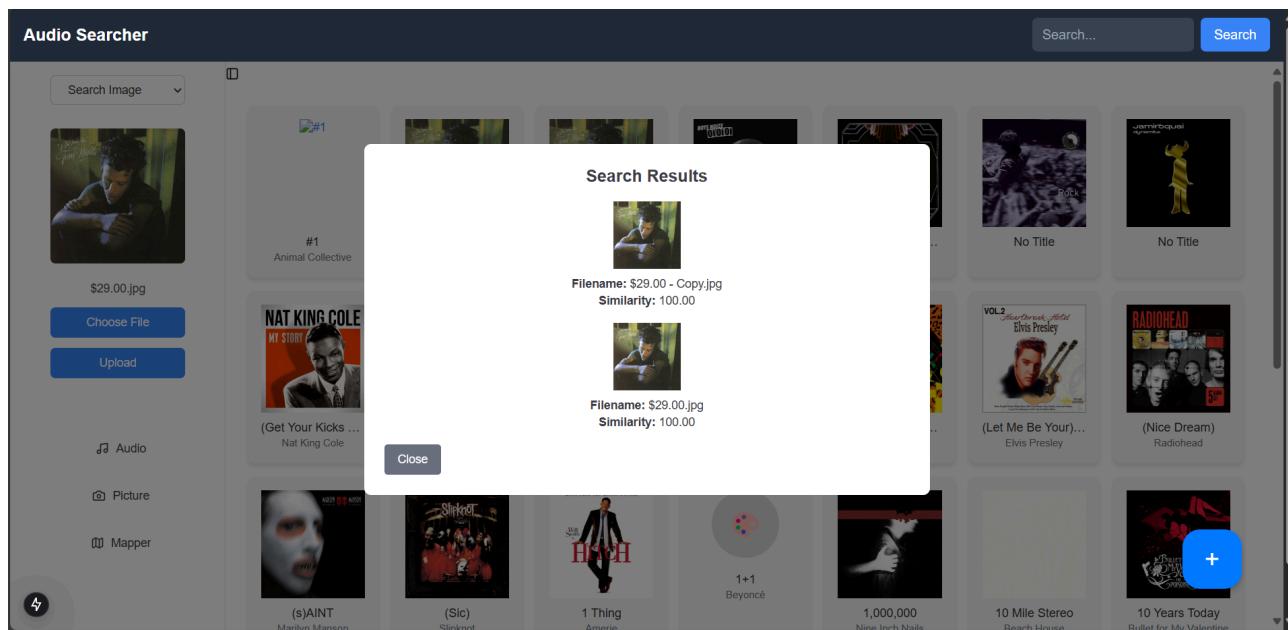
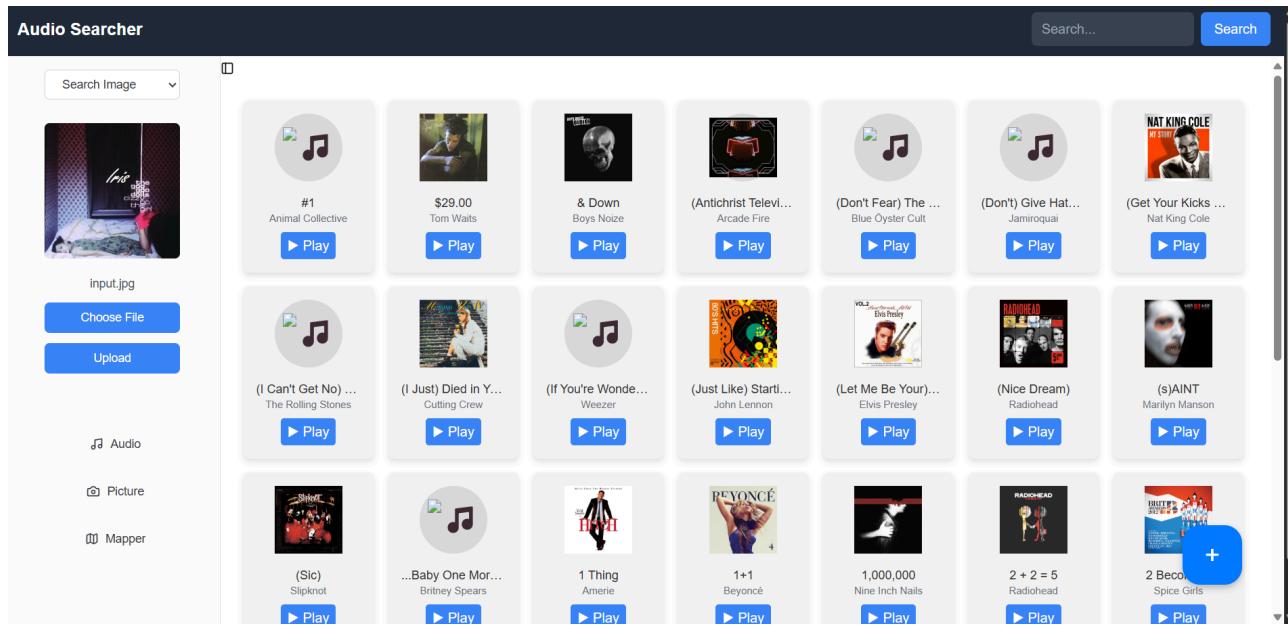
The screenshot shows the "Data Mapper" page of the "Audio Searcher" application. The main content is a table with columns: "No", "index", "thumbnail_image", and "music_preview". The table lists 12 rows of data, corresponding to the items shown in the grid on the previous page. Below the table, there are "Play" buttons for each row and a "close" button. To the right of the table, there are preview thumbnails for "NAT KING COLE", "(s)AINT", and "2 Become 1". At the bottom, there are navigation buttons for "Page 1 of 899" and a blue "+" button.

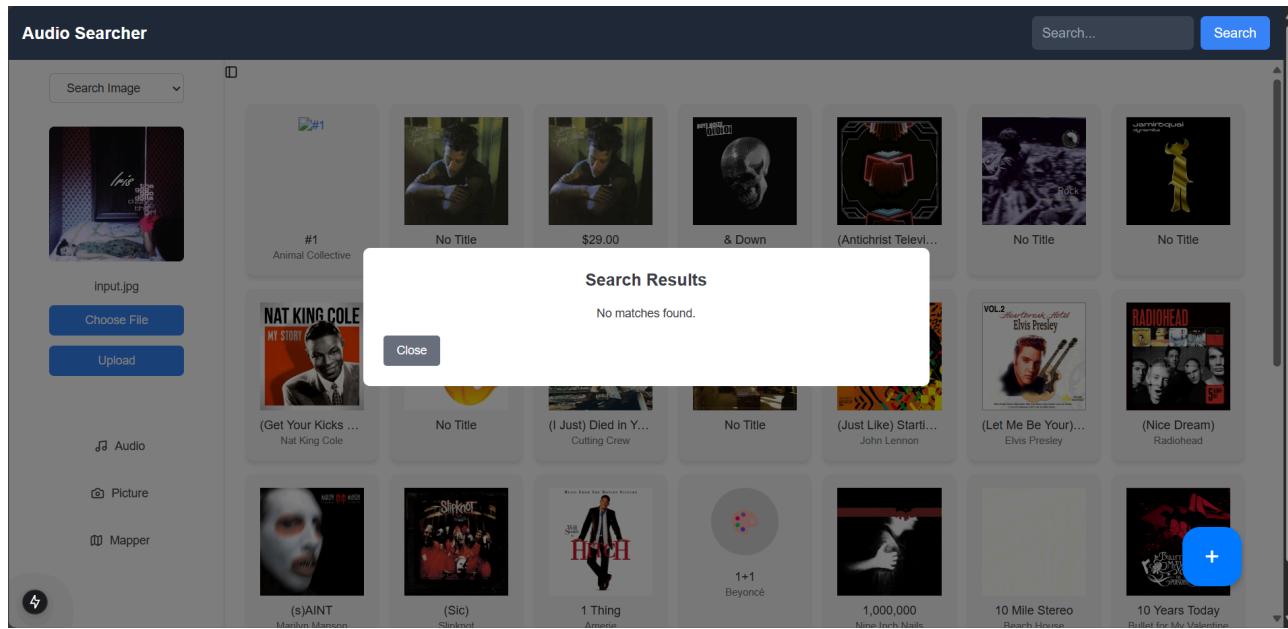
No	index	thumbnail_image	music_preview
1	0	Mr. Brightside.jpg	Mr. Brightside.wav
2	1	Wonderwall.jpg	Wonderwall.wav
3	2	Come as You Are.jpg	Come as You Are.wav
4	3	Take Me Out.jpg	Take Me Out.wav
5	4	Creep.jpg	Creep.wav
6	5	Somebody Told Me.jpg	Somebody Told Me.wav
7	6	Viva la Vida.jpg	Viva la Vida.wav
8	7	Karma Police.jpg	Karma Police.wav
9	8	The Scientist.jpg	The Scientist.wav
10	9	Clocks.jpg	Clocks.wav
11	10	Under the Bridge.jpg	Under the Bridge.wav
12	11	Seven Nation Army.jpg	Seven Nation Army.wav

6. Query by Humming



7. Image Retrieval dengan PCA





BAB V

KESIMPULAN

1. Kesimpulan

Proyek ini berhasil mengimplementasikan *Query by Humming* dan *Image Retrieval* berbasis PCA dengan memanfaatkan antarmuka web. *Query by Humming* memungkinkan pengguna mencari lagu berdasarkan input suara humming. Dengan menerapkan algoritma pencocokan suara, sistem mampu menganalisis pola nada dan mencocokkannya dengan basis data lagu. Sementara itu, *Image Retrieval with PCA* memungkinkan pencarian gambar berdasarkan kesamaan fitur. Sistem ini mengubah gambar menjadi representasi vektor, mengurangi dimensi data menggunakan Principal Component Analysis (PCA), dan mencari gambar yang paling mirip dalam basis data.

Program ini dibangun menggunakan berbagai macam teknologi. Untuk pengolahan gambar, program ini menggunakan bahasa Go. Bahasa Go dipilih karena bahasa ini bisa melakukan sebuah proses dengan paralel sehingga lebih cepat selesai. Untuk pemrosesan suara, program ini menggunakan Python karena kemudahan implementasinya. Untuk *frontend*, program ini memanfaatkan Nextjs untuk membuat tampilan yang interaktif dan mudah digunakan.

2. Komentar

Proses pembuatan program ini merupakan pengalaman yang sangat menarik. Program ini memiliki kegunaan yang lumayan banyak. Akan tetapi, proses pembuatannya cukup rumit. Integrasi algoritma-algoritma ke frontend memakan waktu yang banyak. Apalagi di waktu pengerjaan banyak deadline yang berdekatan, tingkat kesulitannya pun naik dua kali lipat.

3. Saran

Sistem dapat dioptimalkan dengan memanfaatkan caching untuk menyimpan hasil proyeksi gambar dan query yang sering digunakan, sehingga tidak perlu mengulangi proses komputasi yang sama berulang kali. Ini akan mengurangi beban server dan mempercepat waktu respon sistem. Selain itu, juga disarankan untuk melakukan lebih banyak pengujian unit dan pengujian integrasi untuk memastikan bahwa sistem bekerja dengan baik di berbagai kondisi, seperti ketika file input memiliki kualitas buruk atau ukuran yang sangat besar.

LAMPIRAN

Referensi

Marius Kaminskas.2012. *Contextual music information retrieval and recommendation: State of the art and challenges*. Elsevier

[Contextual music information retrieval and recommendation: State of the art and challenges - ScienceDirect](#)

Tautan

Link Github : [salmaanhaniif/Algeo02-23017](#)

Link Video : <https://youtu.be/Whbo00amHgc>