

LAPORAN TUGAS BESAR 2 IF2123

ALJABAR LINIER DAN GEOMETRI

]



Dosen Pengampu :

Ir. Rila Mandala, M.Eng., Ph.D.

Asisten Pembimbing :

Ahmad Naufal Ramadhan

Disusun oleh :

Kelompok 5 (FOTO KELAS)

Rafif Farras 13523095

Barru Adi utomo 13523101

Azfa Radhiyya Hakim 13523115

**PROGRAM STUDI TEKNIK INFORMATIKA
SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA
INSTITUT TEKNOLOGI BANDUNG
JL. GANESA 10, BANDUNG 40132**

2024

DAFTAR ISI

BAB I

DESKRIPSI MASALAH..... **3**

1.1. Tujuan.....	3
1.2. Spesifikasi.....	4

BAB II

TEORI SINGKAT..... **5**

2.1. Image Retrieval System - Principal Component Analysis.....	5
2.1.1. Image Processing and Loading.....	6
2.1.2. Data Centering (Standardization).....	6
2.1.3. PCA Computation Using Singular Value Decomposition (SVD).....	7
2.1.4. Similarity Computation.....	7
2.1.5. Retrieval and Output.....	8
2.2. Music Information Retrieval - Query by Humming.....	9

BAB III

ARSITEKTUR WEBSITE (FRONTEND) DAN ARSITEKTUR PROGRAM

INFORMATION RETRIEVAL (BACKEND)..... **11**

3.1. Arsitektur Website (Frontend).....	11
3.2. Arsitektur Program Information Retrieval (Backend).....	11

BAB IV

EKSPERIMEN..... **13**

4.1. Image Retrieval by Principal Component Analysis.....	13
4.2. Music Retrieval by Query by Humming.....	13

BAB V

KESIMPULAN, SARAN, KOMENTAR DAN REFLEKSI..... **15**

5.1 Kesimpulan.....	15
5.2 Saran.....	15
5.3 Komentar.....	15
5.4 Refleksi.....	16

LAMPIRAN..... **17**

BAB I

DESKRIPSI MASALAH

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 pendekripsi 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.

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. Algoritma yang digunakan adalah *Eigenvalue*, *Cosine Similarity*, *Euclidean Distance*, dll.

1.1. Tujuan

1. Mengimplementasikan sistem Image Retrieval System menggunakan Principal Component Analysis
2. Mengimplementasikan Audio Retrieval System
3. Mengukur similaritas gambar dan audio
4. Menguji kinerja algoritma yang digunakan dalam sistem

1.2. Spesifikasi

1. Buatlah sebuah website yang bisa mendekripsi dan mencari suara dan gambar dalam dataset yang paling cocok

2. Bahasa pemrograman yang digunakan pada frontend dan backend dibebaskan. Mohon dipertimbangkan bahasa-bahasa pemrograman yang kalian gunakan agar bahasa tersebut adalah yang menguntungkan kalian
3. Dataset dari sebuah audio bebas. Silahkan cari di website seperti [kaggle.com](https://www.kaggle.com). Untuk dataset yang akan diujikan, akan diumumkan nantinya.
4. Program dapat memunculkan audio dari dataset ketika sudah diunggah
5. Program dapat mencari kecocokan dari sebuah audio query atau tangkapan dari rekaman audio dengan audio yang ada di dataset berdasarkan metode query by humming
6. Untuk query by humming dapat menggunakan file berformat MIDI ataupun WAV. Silahkan pilih metode manapun yang dapat anda implementasikan.
7. Program dapat memberikan hasil audio / image yang memiliki tingkat kemiripan berkisar antara 55% - 100% (silahkan tentukan sesuai diskusi dari kelompok)
8. Program dapat mengimplementasikan *pagination* untuk menghindari adanya *infinite scrolling* ketika terdapat audio yang terlalu banyak
9. Program dapat menunjukkan jumlah audio yang didapat dari jumlah semua audio dan waktu eksekusi dari program tersebut
10. Bagian bonus hanya boleh dikerjakan apabila spesifikasi wajib dari Tugas Besar telah berhasil dipenuhi. Anda tidak diharuskan untuk mengerjakan keseluruhan bonus, tetapi semakin banyak bonus yang dikerjakan, maka akan semakin banyak tambahan nilai yang diperoleh.
11. Maksimum nilai dari tugas besar ini adalah 115. Biarpun nilai total kalian lebih tinggi karena bonus, nilai kalian akan tetap 115. Silahkan kerjakan bonus jika merasa ingin mengcover nilai spek wajib atau ingin menambah nilai saja.

BAB II

TEORI SINGKAT

2.1. Image Retrieval System - Principal Component Analysis

Principal Component Analysis (PCA) adalah teknik multivariat yang digunakan untuk menganalisis tabel data di mana pengamatan dijelaskan oleh beberapa variabel dependen kuantitatif yang saling berkorelasi. Secara keseluruhan, PCA adalah alat yang kuat untuk menyederhanakan kompleksitas dalam data berdimensi tinggi dan menemukan pola tanpa referensi ke pengetahuan sebelumnya tentang kelompok perlakuan atau perbedaan variabel sebagai titik-titik dalam peta.

PCA sering digunakan dalam berbagai bidang untuk analisis data, termasuk pengurangan dimensi, yang memungkinkan penyederhanaan data berdimensi tinggi sambil mempertahankan tren dan pola yang ada. Teknik ini mengubah data ke dalam dimensi yang lebih sedikit, yang bertindak sebagai ringkasan fitur, dan dapat digunakan untuk mendeteksi outlier serta analisis klaster.

Secara matematis, PCA bergantung pada dekomposisi eigen dari matriks semi-definit positif dan dekomposisi nilai singular (SVD) dari matriks persegi panjang. PCA juga dapat digeneralisasi menjadi analisis korespondensi untuk menangani variabel kualitatif dan analisis faktor ganda untuk menangani set variabel yang heterogen.

Namun, PCA memiliki beberapa keterbatasan, seperti sensitivitas terhadap outlier dan noise, serta tidak adanya metodologi yang jelas untuk kuantifikasi ketidakpastian dari komponen utama atau varians yang dijelaskan. Untuk mengatasi beberapa keterbatasan ini, metode seperti Sparse PCA (SPCA) dan Ensemble PCA (EPCA) telah dikembangkan. SPCA menggunakan lasso untuk menghasilkan komponen utama yang dimodifikasi dengan muatan yang jarang, sementara EPCA menggabungkan PCA yang di-bootstrapped dengan analisis klaster untuk menangani tantangan terkait ambiguitas tanda dan pengurutan ulang komponen dalam sub-sampel PCA.

Secara keseluruhan, PCA adalah alat yang kuat untuk menyederhanakan kompleksitas dalam data berdimensi tinggi dan menemukan pola tanpa referensi ke pengetahuan sebelumnya tentang kelompok perlakuan atau perbedaan fenotipik.

Langkah-langkah untuk melakukan pencarian gambar menggunakan PCA adalah sebagai berikut:

2.1.1. Image Processing and Loading

Lakukan pemrosesan seluruh gambar yang ada pada dataset. Ubah gambar menjadi grayscale untuk mengurangi kompleksitas gambar dan membuat fokus menjadi bagian terang dan gelap gambar. Yang berarti setiap gambar direpresentasikan dalam intensitas pixel 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, gambar akan diubah besarnya sehingga ukurannya sama untuk seluruh gambar. Ukuran seluruh gambar harus konsisten untuk membuat perhitungan semakin akurat. Lalu ubah vektor grayscale pada gambar menjadi 1D untuk membuat dimensi vektor dapat dilakukan pemrosesan data. Jika gambar memiliki dimensi $M \times N$, maka hasilnya adalah vektor dengan panjang $M \cdot N$:

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

2.1.2. Data Centering (Standardization)

Pada step ini lakukan standarisasi data di sekitar nilai 0. Hitung rata rata dari setiap gambar untuk suatu piksel.

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

di mana:

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

Lalu kurangi piksel tersebut dengan rata-rata yang sudah dihitung untuk melakukan standarisasi.

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

2.1.3. PCA Computation Using Singular Value Decomposition (SVD)

Lakukan perhitungan SVD pada gambar-gambar yang sudah distandarisasi.

Buat matriks kovarians dari data yang sudah distandarisasi:

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

di mana:

- X' : matriks data yang sudah distandarisasi.

Lalu lakukan dekomposisi nilai singular untuk mendapatkan komponen utama

$$\mathbf{C} = \mathbf{U}\Sigma\mathbf{U}^T$$

- \mathbf{U} : matriks eigenvector (komponen utama),
- Σ : matriks eigenvalue (menunjukkan varian data di sepanjang komponen utama).

Lalu ambil n jumlah component utama teratas dari hasil SVD dan lakukan proyeksikan gambar ke komponen utama.

Pilih k-komponen utama teratas ($k \ll M \cdot N$) dan proyeksikan data:

$$\mathbf{Z} = \mathbf{X}'\mathbf{U}_k$$

di mana:

- \mathbf{U}_k : matriks eigenvector dengan n-dimensi.

2.1.4. Similarity Computation

Lakukan pencarian kesamaan dengan mencari jarak euclidean dari tiap dataset dengan gambar query. Lalu lakukan pengurutan kecocokan dari yang paling tinggi.

Pertama-tama, representasikan gambar query dalam ruang komponen utama dengan proyeksi yang sama:

$$\mathbf{q} = (\mathbf{q}' - \mu)\mathbf{U}_k$$

Dimana:

- \mathbf{q} = Vektor proyeksi dari gambar query ke ruang komponen utama (PCA).
- \mathbf{q}' : Gambar query dalam format vektor (setelah grayscale, resize, dan flattening).
- μ : Rata-rata piksel dari dataset (per piksel).
- \mathbf{U}_k : Matriks eigenvector dengan k dimensi utama dari PCA.

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

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

- $d(\mathbf{q}, \mathbf{z}_i)$ = Jarak antara gambar query \mathbf{q} dan gambar ke- i dalam ruang komponen utama.
- \mathbf{z}_i = Vektor proyeksi dari gambar ke- i dalam dataset ke ruang komponen utama.
- q_j : Elemen ke- j dari vektor proyeksi query \mathbf{q} .
- z_{ij} : Elemen ke- j dari vektor proyeksi gambar ke- i , yaitu \mathbf{z}_i .
- k : Jumlah dimensi ruang komponen utama yang dipilih.

Lalu, Urutkan hasil berdasarkan jarak terkecil.

2.1.5. Retrieval and Output

Kumpulkan gambar-gambar yang mirip dengan query masukan dengan cara melakukan limitasi jumlah atau dengan memberikan batas jarak euclidean.

Hasil dari pencarian gambar dapat digabungkan dengan hasil pencarian suara ataupun dijalankan sendiri.

2.2. Music Information Retrieval - Query by Humming

Query by humming adalah suatu teknik pencarian musik yang memanfaatkan melodi yang didengungkan atau dinyanyikan oleh pengguna untuk dicocokkan dengan database lagu yang tersedia. Sistem ini menganalisis pola nada dari input pengguna, mengubahnya menjadi representasi numerik (pitch contour), dan membandingkannya dengan representasi serupa dalam database.

Terdapat beberapa langkah utama dalam menjalankan algoritma query ini. Pertama, dilakukan pemrosesan audio, yang didapat dari rekaman suara pengguna ataupun masukan file, yang kemudian akan dimanfaatkan nada-nada yang tersedia. Terdapat dua macam file yang dapat digunakan, yaitu wav file dan midi file. Kami menggunakan library librosa untuk mengekstrak file wav dan mido untuk mengekstrak file midi. Untuk file midi, secara umum akan digunakan nada pada channel 1. Namun ada kalanya, terdapat file midi yang tidak memiliki channel rendah. Oleh karena itu, kami akan mencari ke nada pada channel selanjutnya (secara traversal) hingga minimal didapatkan 30 panjang nada. Selanjutnya, hasil dari pengekstrakan file tersebut akan dibagi menjadi 30 segmen dan 6 beat sliding window. Hal ini kami lakukan agar mendapatkan hasil kemiripan yang seimbang dengan efisiensi waktu dan memori. Untuk menjaga kestabilitasan data, nada-nada tersebut akan dinormalisasi tempo dengan membagi selisih note (untuk setiap note) dengan rata-rata dari pitch, dengan standar deviasi.

$$NP(note) = \frac{(note-\mu)}{\sigma}$$

Kemudian, data tersebut akan diekstraksi berdasarkan tiga fitur, yaitu Absolute Tone Based (ATB), Relative Tone Based (RTB), dan First Tone Based (FTB). Data-data tersebut kemudian akan kembali di normalisasi agar semua nilai histogram berada dalam skala probabilitas.

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

Langkah berikutnya, untuk membandingkan suatu file dengan file yang lain, akan dibandingkan nilai kedua vektor menggunakan cosine similarity. Semakin besar nilainya (mendekati 1), maka kemiripan kedua file tersebut semakin besar. Dalam menghitung kemiripan tersebut, akan digunakan perbandingan ATB, RTB, dan FTB dengan perbandingan 5 : 3 : 2. Hal tersebut didasari fakta bahwa ATB mengandung fitur yang lebih dominan dibandingkan fitur lain nya. Selain itu, RTB memiliki pengaruh yang relatif lebih tinggi dibandingkan FTB, yang memberikan informasi tambahan yang mendukung, namun tidak dominan.

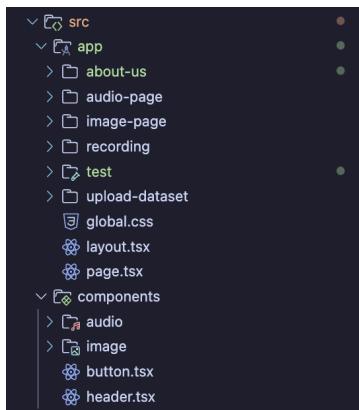
$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

BAB III

ARSITEKTUR WEBSITE (FRONTEND) DAN ARSITEKTUR PROGRAM INFORMATION RETRIEVAL (BACKEND)

3.1. Arsitektur Website (Frontend)

Pada front-end, website menggunakan techstack Next.js, Tailwind CSS untuk styling, dan Axios penghubung front-end dan back-end. Pada folder App, terdapat page yang dapat diakses di

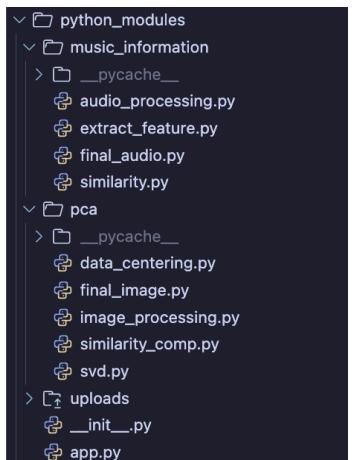


route website, diantaranya yaitu about-us, audio-page, image-page, dan upload-dataset. Terdapat juga folder components untuk memasukkan component yang dipakai berulang kali, seperti AudioPlayCard, AudioCard, AudioSubmit, ImageList, ImageCard, dan Header.

Pada frontend, kami menggunakan TypeScript agar lebih aman dalam mendeklarasikan tipe data dan meningkatkan keterbacaan kode.

3.2. Arsitektur Program Information Retrieval (Backend)

Pada backend, website menggunakan tech stack FastAPI sebagai penghubung front-end dan back-end, python untuk memproses datanya. Pada pemrosesan audio, library python yang digunakan adalah numpy, mido, dan librosa sedangkan pada pemrosesan gambar, kami menggunakan library numpy.



Beberapa api yang digunakan adalah:

- **/mapper**, yang digunakan untuk mengambil data mapper
- **/dataset**, yang digunakan untuk mengambil data dari folder audio jika tidak ada mapper
- **/upload_file, /upload_audio, /upload_image**, yang digunakan untuk upload file dari front-end ke back-end.
- **/compare/audio** dan **/compare/image** untuk mengirimkan hasil komparasi audio dan image dengan database.

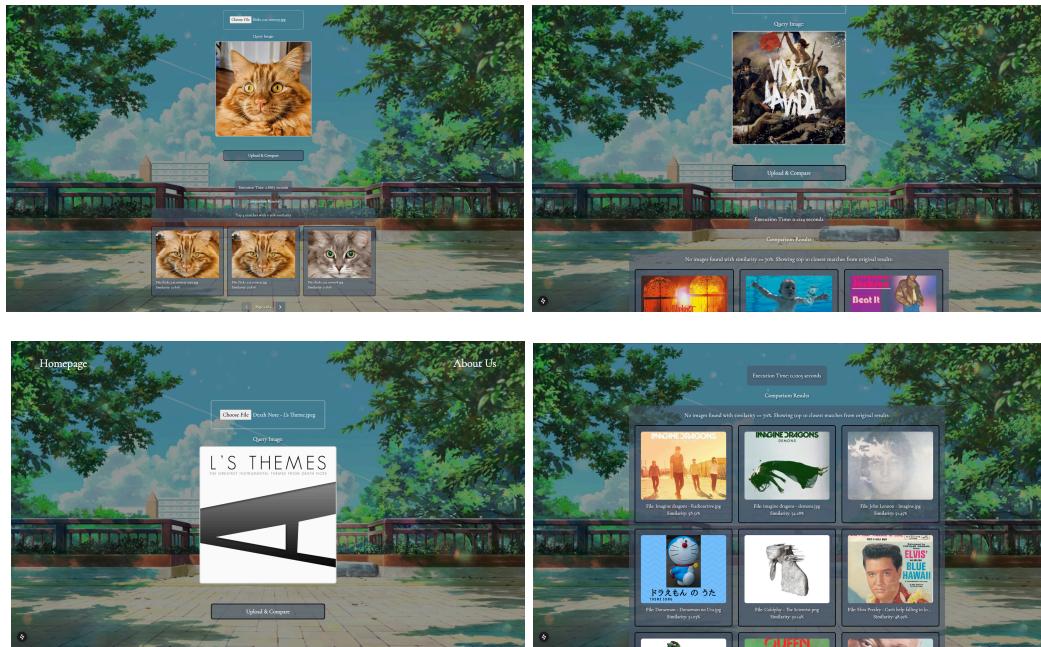
Dalam mengakses database, kami menggunakan sistem *cache* untuk mempercepat proses perhitungan secara berkala. *Cache* ini kami gunakan ketika pengguna mencoba untuk membandingkan gambar/audio untuk pertama kalinya. Hasil dari pemrosesan tersebut kami manfaatkan agar dapat digunakan pada perhitungan selanjutnya (jika memakai database yang sama). Pada page upload-dataset, kami menyediakan fitur clear database untuk menghilangkan cache yang telah disimpan.

BAB IV

EKSPERIMENT

4.1. Image Retrieval by Principal Component Analysis

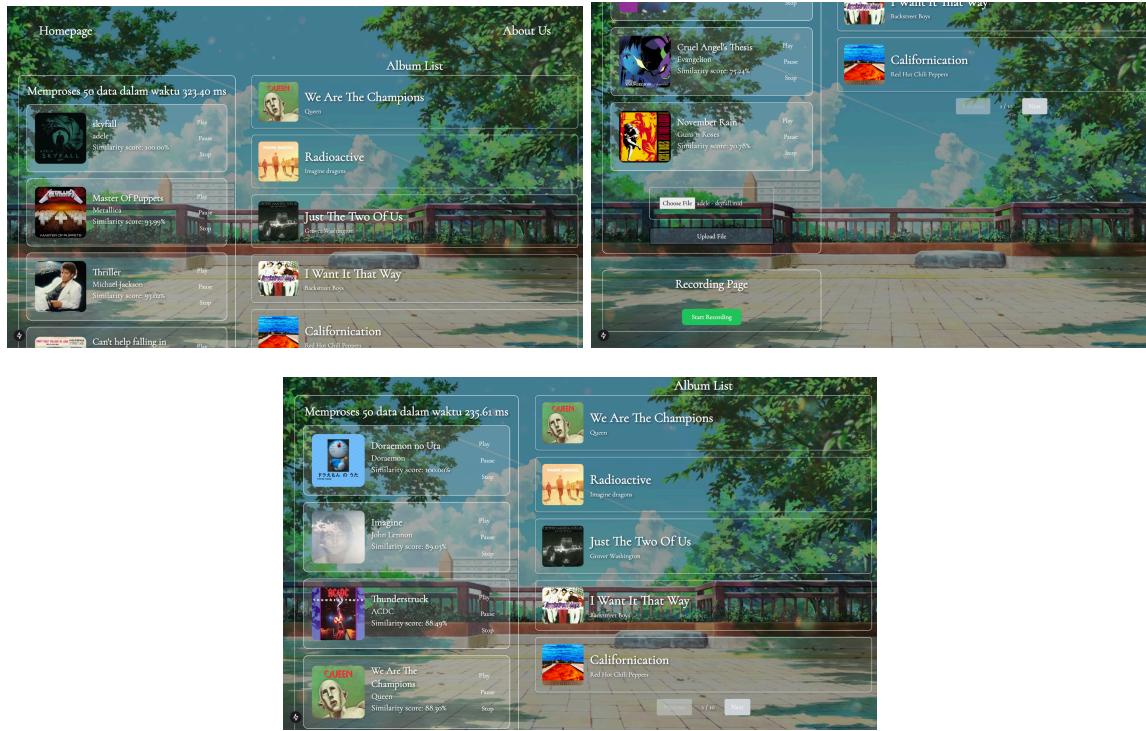
Berikut merupakan eksperimen untuk image retrieval dengan PCA.



Setelah menekan submit, program kemudian akan menampilkan semua gambar yang memiliki persentase kemiripan lebih dari 70%. Jika tidak ada hasil yang memenuhi batas 70%, maka akan menampilkan 10 image dengan hasil komparasi terbesar.

4.2. Music Retrieval by Query by Humming

Pada bagian kanan music retrieval, terdapat list album yang tersedia sebagai dataset dan pada bagian kiri terdapat input dengan file maupun recording. Ketika pengguna menekan tombol record, maka program akan secara otomatis merekam suara yang ada disekitar, dan akan secara otomatis berhenti setelah 5 detik jika data yang didapat sudah dianggap cukup. Jika proses telah selesai, akan ditampilkan hasil kemiripan dengan skor lebih dari 70%.



4.3 Upload Dataset

Pada upload dataset, terdapat 3 bagian yaitu upload image, upload audio, dan upload mapper. Kemudian dapat di-submit dengan tombol submit atau dapat di-delete dataset yang sudah ada dengan tombol delete.



BAB V

KESIMPULAN, SARAN, KOMENTAR DAN REFLEKSI

5.1 Kesimpulan

Tugas Besar 2 Aljabar Linier dan Geometri ini berfokus pada implementasi sistem temu balik suara dan gambar. Melalui penerapan metode Principal Component Analysis (PCA) untuk gambar dan teknik humming untuk audio, tugas ini memberikan kontribusi signifikan kepada diri sendiri dan dalam bidang Information Retrieval. Penulis berhasil mengimplementasikan metode ekstraksi fitur yang komprehensif menggunakan pendekatan Principal Component Analysis (PCA) untuk gambar dan teknik humming untuk audio. Sistem yang dikembangkan mampu mentransformasi data menjadi representasi numerik yang dapat diproses secara efisien, dengan menggunakan algoritma seperti *Cosine Similarity* dan *Euclidean Distance* untuk mengukur tingkat kemiripan. Integrasi metode ekstraksi fitur dari kedua domain-audio dan visual memungkinkan pencarian informasi yang lebih mendalam dan kontekstual, melampaui keterbatasan pencarian berbasis kata kunci tradisional.

Selama proses pengembangan, penulis menghadapi sejumlah tantangan kompleks, mulai dari permasalahan teknis ekstraksi fitur hingga optimasi algoritma pencocokan. Kompleksitas dalam menangkap karakteristik unik dari setiap jenis media, menyesuaikan variasi kualitas input, dan mencapai akurasi tinggi dengan waktu sesingkat mungkin merupakan beberapa kendala utama yang memerlukan pendekatan inovatif dan berkelanjutan.

5.2 Saran

Saran untuk tugas besar ini adalah pemberian spesifikasi yang bisa lebih diperjelas dan pemberian waktu yang lebih lama dikarenakan masih dalam tahap belajar membuat website dengan framework baru, ditambah lagi dengan adanya spesifikasi tertentu seperti waktu yang membuat pemrosesan pada backend harus dibuat seoptimal mungkin.

5.3 Komentar

Tugas besar kedua ini sangat bermanfaat bagi penulis dapat mempelajari bagaimana konsep-konsep matematis tersebut diaplikasikan dalam konteks sistem temu balik gambar dan

audio, yang merupakan salah satu teknologi penting dalam pemrosesan citra dan sinyal. Implementasi ini membuka wawasan baru bagi penulis mengenai bagaimana teori yang dipelajari di bangku kuliah dapat diadaptasi untuk menyelesaikan masalah nyata di dunia industri, seperti pencarian gambar berdasarkan fitur atau pengenalan pola dalam audio. Kemudian dengan kolaborasi kelompok yang baik, tugas besar ini dapat diselesaikan dengan baik.

5.4 Refleksi

Dari pengalaman Tugas Besar Algeo 2 ini, ada beberapa kekurangan pada kelompok penulis, seperti lama waktu yang dibutuhkan dalam membandingkan image dengan dataset dan juga dalam membandingkan audio dengan dataset, namun hal ini dapat diperbaiki kedepannya untuk meningkatkan kualitas output yang dihasilkan. Penulis juga menyadari perlunya perbaikan pada proses query by humming agar mendapat hasil yang lebih akurat. Selain itu, penggunaan dataset yang lebih beragam dapat meningkatkan validitas hasil dan membantu mengidentifikasi potensi kelemahan dalam algoritma. Refleksi ini akan menjadi landasan untuk perbaikan di masa mendatang yang mengarahkan penulis untuk terus meningkatkan kualitas dan skill khususnya dalam pengembangan proyek yang berbasis aljabar geometri.

LAMPIRAN

Dorabiala, O., Aravkin, A., & Member, I. (2023). Ensemble Principal Component Analysis. *IEEE Access*, 12, 6663-6671. <https://doi.org/10.1109/ACCESS.2024.3350984>.

Gemperline, P. (2003). Principal Component Analysis. *Technometrics*, 45, 276 - 276. https://doi.org/10.1007/978-0-387-30164-8_665.

Whitlark, D., & Dunteman, G. (1990). Principal Components Analysis. *Journal of Marketing Research*, 27, 243. <https://doi.org/10.2307/3172855>.

Bu, Y., Wong, R., & Fu, A. (2009). Query by Humming. , 2244-2249. https://doi.org/10.1007/978-0-387-39940-9_292.

Foote, J. (1999). An overview of audio information retrieval. *Multimedia Systems*, 7, 2-10. <https://doi.org/10.1007/s005300050106>.

Tautan Repository:

<https://github.com/azfaradhi/Algeo02-23095>

Tautan Video:

<https://youtu.be/braReAxdIGU?feature=shared>