

# **MUSIC RECOGNITION USING DEEP LEARNING**

## **PROPOSAL UBICOM**



**uin**  
UNIVERSITAS ISLAM NEGERI  
SUNAN GUNUNG DJATI  
BANDUNG

**Oleh**

**Muhammad Rahmat Muhaimin**

**1207050078**

**TEKNIK INFORMATIKA  
SAINS DAN TEKNOLOGI  
UIN SUNAN GUNUNG DJATI BANDUNG  
BANDUNG  
2023M/1444H**

# MUSIC RECOGNITION USING DEEP LEARNING

## 1.1. Latar Belakang Penelitian

Dalam zaman modern, industri musik telah mengalami pertumbuhan signifikan dan telah menjadi bagian integral dari kehidupan sehari-hari banyak individu[1]. Para penggemar musik mendengarkan berbagai lagu yang diciptakan oleh berbagai artis secara teratur. Meskipun mudah untuk menemukan lagu sesuai selera kita melalui pencarian online, menemukan lagu baru di tempat umum merupakan tantangan yang berbeda. Mendengar lagu untuk pertama kalinya di tempat umum membuat sulit untuk mengidentifikasi judul lagu, sehingga menjadi tidak mungkin untuk mendengarkannya lagi[2]. Misalnya, seseorang mungkin mendengar lagu yang sedang diputar di latar belakang saat menunggu di halte bus, tetapi tanpa pengetahuan tentang judul atau artisnya, menjadi tidak mungkin untuk menemukan lagu tersebut lagi. Bagaimana jika ada aplikasi yang dapat menangkap potongan lagu yang sedang diputar dan mengembalikan nama lagu beserta URL-nya? Itu akan sangat bagus, oleh karena itu kami mengembangkan proyek ini yang dapat mengenali potongan audio pendek yang sedang diputar.

Ada beberapa aplikasi pengenalan musik yang tersedia di pasaran, seperti Shazam[3], SoundHound[4], dan Google Sound Search[5], yang mampu mengidentifikasi lagu. Setiap aplikasi menggunakan algoritma yang unik untuk pengenalan lagu. Dalam proyek saya, kami mengevaluasi berbagai algoritma yang digunakan dalam bidang pengenalan musik dan mengidentifikasi yang paling cocok untuk aplikasi kami. Proses pembangunan sistem dimulai dengan mengonversi audio ke format yang diinginkan menggunakan teknik pemrosesan sinyal. Selanjutnya, audio diubah menjadi log mel-spectrogram[6], yang merupakan representasi waktu, frekuensi, dan kebisingan suara menggunakan Short Time Fourier Transform (STFT), mel filter bank, dan konversi skala dB. Log mel-spectrogram berfungsi sebagai fitur utama audio. Untuk meningkatkan akurasi model, versi audio yang telah diubah dibuat dengan menggabungkan noise latar belakang, IR impulse response, dan teknik lainnya. Log mel-spectrogram dihasilkan untuk setiap klip audio yang telah diubah, dan kedua

versi asli dan yang telah diubah digunakan untuk pelatihan jaringan saraf. Pendekatan pembelajaran kontrastif digunakan selama proses pelatihan untuk memastikan bahwa embedding lagu asli dan versi yang telah diubah berdekatan, sementara embedding lagu lain berada lebih jauh. Model dioptimalkan dengan meminimalkan fungsi kerugian dalam setiap epoch[7]. Setelah model dilatih, digunakan untuk menghasilkan sidik jari untuk setiap lagu, yang disimpan dalam database. Sidik jari ini dibandingkan dengan sidik jari potongan audio input menggunakan teknologi FAISS[8].

Untuk membuat sistem ini dapat diakses oleh pengguna, model ini diintegrasikan ke dalam aplikasi seluler yang menyediakan antarmuka interaktif. Mikrofon seluler digunakan untuk menangkap audio, dan model yang telah dilatih digunakan untuk menghasilkan sidik jari audio yang ditangkap. Sidik jari dibandingkan dengan sidik jari dalam database, dan nama, URL, dan informasi lainnya tentang lagu yang cocok dikembalikan kepada pengguna.

## 1.2. Branding

Pada Tahap ini saya menentukan branding dari produk UbiCom yang dibuat, antara lain:

1. Merk : MusiScan
2. Inspirasi Merk : Nama ini menggabungkan unsur "Musik" yang berkaitan dengan fungsinya dan "Scan" yang mencerminkan kemampuannya untuk "menggali" potongan audio dan mengidentifikasi lagu.
3. Tagline : Pencari Judul Lagu
4. Campaign : Bagaimana Menjadikan *Handphone* yang dimiliki sebagai perangkat yang bisa membantu mencari lagu yang ingin diketahui di tempat umum
5. Target User :
  - Usia 7+
  - Memiliki Ketertarikan dengan Musik
  - Musisi
6. User Experience Theme :
  - Mudah dikonfigurasi dan digunakan

- Hanya menggunakan handphone
- Hanya membutuhkan internet

### 1.3. User Story

Pada tahap ini saya mengeksplorasi kebutuhan prioritas dari para pengguna untuk diwujudkan sebagai fitur pada sistem atau aplikasi yang akan dibuat.

No.	Sebagai	Saya Ingin Bisa	Sehingga	Prioritas
1	Pengguna musik yang ingin menemukan lagu baru dari audio yang sedang diputar	merekam audio sekitar saya dan mengidentifikasi judul lagu, artis, dan penyedia tautan ke lagu tersebut.	saya dapat dengan mudah menemukan dan mendengarkan lagi lagu yang saya suka tanpa harus mencari tahu secara manual.	Tinggi
2	pengguna yang suka menggabungkan lagu-lagu yang baru saya temukan ke dalam daftar putar pribadi,	menambahkan lagu yang diidentifikasi ke dalam daftar putar favorit saya.	saya dapat menyusun daftar putar saya dengan mudah dan mendengarkan kembali lagu-lagu favorit saya kapan saja.	Sedang
3	pengguna yang ingin mendapatkan rekomendasi lagu berdasarkan preferensi saya	melihat rekomendasi lagu yang mirip dengan lagu yang telah saya identifikasi.	saya dapat mengeksplorasi lebih banyak musik yang sesuai dengan selera saya.	Rendah

#### 1.4. Metode dan Algoritma

Pada Tahap Ini saya akan menjelaskan metode dan algoritma apa yang digunakan pada setiap komponen aplikasi yang dibuat, antara lain:

1. Sensor:

- Audio Recognition: Penggunaan mikrofon perangkat seluler untuk merekam audio dari lingkungan sekitar. Disini akan memerlukan algoritma pemrosesan audio yang mampu mengubah suara menjadi data yang dapat dianalisis. Ini melibatkan teknik pengambilan sampel, analisis spektrum frekuensi, dan pencocokan pola audio untuk mengenali lagu.

2. Audio Analysis: Analisis audio melibatkan pemrosesan data suara, seperti pengukuran spektrum frekuensi, amplitudo, dan waktu, untuk mengidentifikasi fitur-fitur audio yang akan digunakan dalam pengenalan lagu.

3. Contrastive Learning: Metode pembelajaran kontrastif digunakan untuk melatih model agar memahami perbedaan antara lagu yang diidentifikasi dan lagu lainnya. Ini melibatkan penggunaan teknik-teknik seperti:

- Convolutional encoder: Penggunaan encoder konvolusional untuk mengubah data audio menjadi representasi yang lebih informatif.
- Divide and Encode Layer: Lapisan yang membagi dan mengkodekan data audio ke dalam representasi yang lebih kecil.
- NTxent: Algoritma yang digunakan dalam pelatihan kontrastif untuk memastikan bahwa embedding lagu yang sama mendekati satu sama lain sementara embedding lagu yang berbeda berjauhan

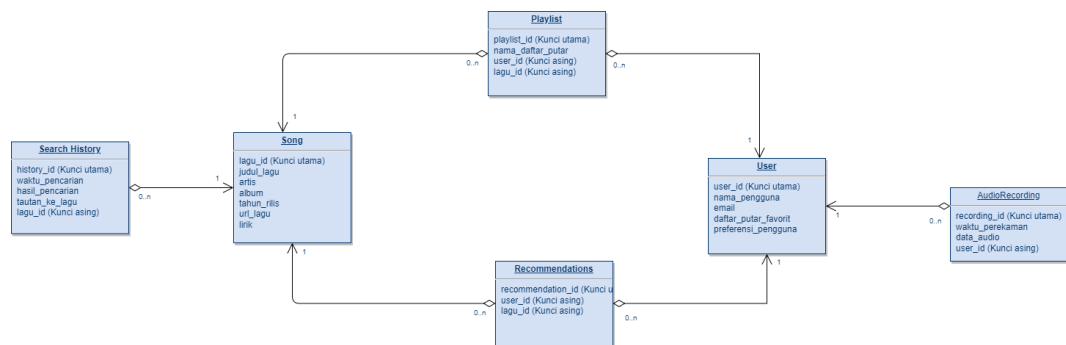
4. Optimizer: Optimizer digunakan dalam pelatihan model untuk mengoptimalkan parameter. Dua jenis optimizer yang disebutkan adalah:

- ADAM Optimizer: Salah satu metode optimasi yang digunakan untuk memperbarui parameter model selama pelatihan.

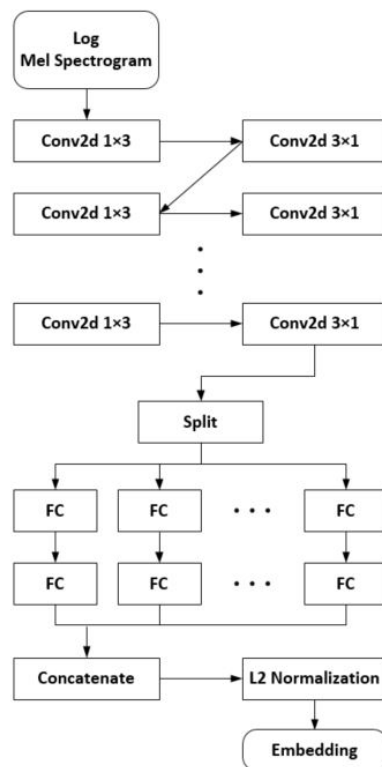
- Lamb Optimizer: Jenis lain dari optimizer yang dapat digunakan untuk meningkatkan konvergensi model.
5. Sequence Search: Sequence search mengacu pada kemampuan aplikasi untuk mencari dan mencocokkan urutan audio dengan lagu dalam database berdasarkan fitur-fitur yang dihasilkan selama analisis audio dan pelatihan model.

## 1.5. Struktur Data

Pada Tahap ini saya mengeksplorasi struktur data yang digunakan pada aplikasi yang dibuat, antara lain:



## 1.6. Arsitektur Model



**Gambar 1.1. Arsitektur Model**

## 1.7. Deskripsi Teknologi

Pada tahap ini saya akan menjelaskan setiap teknologi hardware dan software yang digunakan dalam Pembangunan sistem, antara lain:

### 1. Python

Dalam proyek kami, kami menggunakan bahasa pemrograman Python untuk membangun sistem. Python adalah bahasa pemrograman yang serbaguna dan kuat yang memiliki komunitas pengembang yang besar dan aktif yang berkontribusi pada pengembangan, dokumentasi, dan dukungan. Komunitas seperti ini juga dengan cara tertentu membantu proyek kami. Alasan utama memilih Python untuk proyek kami adalah karena fitur-fitur kuncinya seperti kemudahan dibaca, kesederhanaan, dan perpustakaan standar yang besar yang menyediakan pengembang dengan beragam alat dan fungsi, serta jumlah perpustakaan dan kerangka kerja pihak ketiga yang besar dan terus berkembang.

### 2. TensorFlow

TensorFlow adalah perpustakaan perangkat lunak sumber terbuka untuk pemrograman dataflow dan berbeda pada berbagai tugas, termasuk pembelajaran mesin, pembelajaran mendalam, dan jaringan saraf. Ini memungkinkan peneliti dan pengembang untuk dengan cepat dan mudah membangun dan melatih model pembelajaran mesin. Kami menggunakan perpustakaan ini dalam proyek kami untuk membangun dan melatih model jaringan saraf kami yang terdiri dari berbagai lapisan.

### 3. Kapre

Kapre adalah perpustakaan perangkat lunak sumber terbuka untuk pemrosesan audio dengan jaringan saraf mendalam. Kami telah menggunakan perpustakaan ini dalam proyek kami untuk pemrosesan audio dan menghasilkan log mel-spectrogram (fitur audio) dari audio mentah. Berbagai fungsi bawaan seperti STFT, Magnitude, ApplyFilterBank, dll. diimpor dari perpustakaan ini dalam proyek kami.

### 4. Flutter

Dalam proyek kami, kami menggunakan Flutter untuk pengembangan aplikasi seluler, yang merupakan kerangka

pengembangan aplikasi seluler sumber terbuka yang menyediakan cara yang kuat dan efisien untuk membuat aplikasi seluler lintas platform yang dapat berjalan baik di platform Android maupun iOS menggunakan satu kode sumber. Ini didasarkan pada bahasa pemrograman Dart. Fitur-fitur seperti lintas platform, pengembangan cepat, kinerja tinggi, antarmuka pengguna yang indah, dan komunitas aktif menarik kami untuk menggunakan Flutter dalam proyek kami.

#### 5. Google Colab

Google Colab adalah platform berbasis awan untuk mengembangkan dan menjalankan kode Python dalam peramban web. Dalam proyek kami, kami membangun dan melatih model kami di Google Colab. Lingkungannya yang telah dikonfigurasi sebelumnya dengan perpustakaan ilmu data populer dan akses ke perangkat keras kuat seperti GPU dan TPU membuatnya mudah untuk membangun model kami dan mempercepat pelatihan model. Yang terbaik adalah kita dapat menggunakannya secara gratis, yang sangat berharga untuk menyelesaikan proyek dengan anggaran kecil.

#### 6. Microsoft Visio

Dalam proyek kami, kami menggunakan Microsoft Visio untuk membuat semua hal berbentuk diagram yang dibutuhkan dalam proyek kami. Dengan menggunakan templat dan berbagai bentuk yang disediakan oleh Microsoft Visio, kami menggambar arsitektur sistem, diagram use case, diagram aktivitas, dan diagram urutan dengan efektif.

#### 7. yt-dlp dan ffmpeg

Yt-dlp adalah program/alat baris perintah sumber terbuka yang telah kami gunakan dalam proyek kami untuk mengunduh audio, thumbnail, dan metadata video musik di YouTube. FFmpeg adalah perangkat lunak sumber terbuka lintas platform yang digunakan untuk memproses dan mengonversi file audio dan video. Kami menggunakannya dalam proyek kami untuk mengonversi format audio .opus menjadi format .wav dan juga mengubah tingkat sampling menjadi 8000Hz dengan saluran mono.



## 8. Visual Studio Code

Visual Studio Code (VS Code) adalah editor kode gratis dan sumber terbuka yang dikembangkan oleh Microsoft. Fitur-fitur seperti ringan, cepat, sangat dapat disesuaikan, dan ketersediaan ekstensi/plugin telah membuat kita lebih mudah dalam pemrograman.

## 9. FAISS

FAISS (Facebook AI Similarity Search) adalah perpustakaan sumber terbuka yang dikembangkan oleh tim Penelitian Kecerdasan Buatan Facebook untuk pencarian kesamaan dan pengelompokan vektor berdimensi tinggi dengan efisien. Ini digunakan dalam proyek kami karena kemudahannya dan kemampuannya untuk melakukan pencarian tetangga terdekat dengan cepat dan efisien dalam kumpulan data besar yang berisi jutaan atau miliaran vektor berdimensi tinggi.

### 1.8. User Experience (UX) Design

Pada Tahap ini saya akan menjelaskan fitur Mobile App Yang kemungkinan akan dibuat, antara lain:

1. User Authentication: Sistem ini memungkinkan pengguna untuk mendaftar akun baru atau mengakses yang sudah ada dengan menyediakan alamat email atau kredensial masuk media sosial. Selain itu, pengguna memiliki opsi untuk mengatur ulang kata sandi jika mereka lupa.
2. Audio Record: Dengan fitur perekaman audio dalam aplikasi, pengguna dapat mencari detail tentang lagu. Mereka dapat merekam audio di dalam aplikasi dan menggunakannya untuk mengidentifikasi lagu melalui fungsi pencarian.
3. Audio Upload: Aplikasi memungkinkan pengguna untuk mengunggah file audio dari perangkat mereka, yang dapat digunakan untuk pencarian atau disimpan dalam profil mereka untuk referensi di masa depan.
4. History: Setelah mencari lagu, aplikasi menyimpan hasil pencarian dalam riwayat pengguna, yang dapat diakses dan dicari kapan saja oleh pengguna.

5. Songs Identification: Fungsi utama dari aplikasi seluler ini adalah mengidentifikasi lagu berdasarkan klip audio yang diajukan oleh pengguna.
6. Interactive Interface: Antarmuka aplikasi telah dirancang untuk mudah digunakan, memungkinkan navigasi dan pemahaman yang mudah oleh semua pengguna.
7. Recommend Songs: Selain itu, aplikasi juga menyarankan lagu-lagu dengan genre serupa dengan hasil pencarian serta lagu-lagu lain oleh artis yang sama.

## DAFTAR PUSTAKA

- [1] P. K. Khanal, A. Kayastha, and A. Khatakho, "MUSIC RECOGNITION USING DEEP LEARNING," 2022, Accessed: Oct. 21, 2023. [Online]. Available: <https://elibrary.tucl.edu.np/handle/123456789/18844>
- [2] "pradeep k khnal et al. be report electronics apr 2022".
- [3] Shazam, "Shazam," <https://www.shazam.com/>.
- [4] SoundHound Inc, "SoundHound ," <https://www.soundhound.com/>.
- [5] Google, "Google assistant," <https://assistant.google.com/services/a/uid/00000024216d4bb8?hl=en-US>.
- [6] J. Johnson, M. Douze, H. J.-I. T. on B. Data, and undefined 2019, "Billion-scale similarity search with gpus," *ieeexplore.ieee.org* J Johnson, M Douze, H Jégou *IEEE Transactions on Big Data*, 2019 • *ieeexplore.ieee.org*, Accessed: Oct. 21, 2023. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8733051/>
- [7] H. Jégou, M. Douze, and C. Schmid, "Product quantization for nearest neighbor search," *IEEE Trans Pattern Anal Mach Intell*, vol. 33, no. 1, pp. 117–128, 2011, doi: 10.1109/TPAMI.2010.57.
- [8] J. F. Gemmeke *et al.*, "Audio set: An ontology and human-labeled dataset for audio events," *ieeexplore.ieee.org* JF Gemmeke, DPW Ellis, D Freedman, A Jansen, W Lawrence, RC Moore, M Plakal, M Ritter 2017 *IEEE international conference on acoustics, speech and signal*, 2017 • *ieeexplore.ieee.org*, Accessed: Oct. 21, 2023. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/7952261/>