

**KLASIFIKASI GENRE MUSIK POP ISLAMI
MENGUNAKAN CONVOLUTIONAL NEURAL NETWORK
(CNN)**

LAPORAN TUGAS AKHIR

Proposal ini disusun guna memenuhi salah satu syarat untuk menyelesaikan
program studi Teknik Informatika S-1 pada Fakultas Teknologi Industri
Universitas Islam Sultan Agung



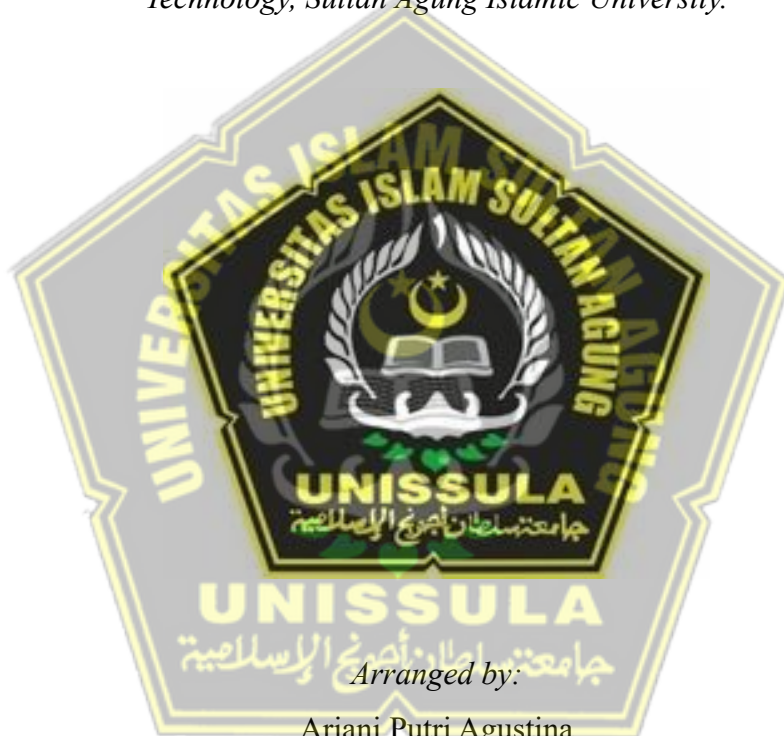
Disusun Oleh:
Ariani Putri Agustina
32602100033

**FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM SULTAN AGUNG
SEMARANG
2025**

CLASSIFICATION OF ISLAMIC POP MUSIC GENRES USING CONVOLUTIONAL NEURAL NETWORK (CNN)

FINAL PROJECT

This report was prepared to fulfill one of the requirements for completing the Undergraduate Informatics Engineering Study Program at the Faculty of Industrial Technology, Sultan Agung Islamic University.



Arranged by:

Ariani Putri Agustina

32602100033

***MAJORING OF INFORMATICS ENGINEERING
FACULTY OF INDUSTRIAL TECHNOLOGY
SULTAN AGUNG ISLAMIC UNIVERSITY***

2025

**LEMBAR PENGESAHAN
TUGAS AKHIR**

**KLASIFIKASI GENRE MUSIK POP ISLAMI MENGGUNAKAN
CONVOLUTIONAL NEURAL NETWORK (CNN)**

**ARIANI PUTRI AGUSTINA
NIM 32602100033**

Telah dipertahankan di depan tim penguji ujian sarjana tugas akhir
Program Studi Teknik Informatika
Universitas Islam Sultan Agung
Pada tanggal : 12 Juni 2025

TIM PENGUJI UJIAN SARJANA :

Andi Riansyah, ST., M.Kom
NIK. 210616053
(Ketua Penguji)

26-06-2025

Bagus Satrio WP, S.Kom., M.Cs
NIK. 210616051
(Anggota Penguji)

25-06-2025

Imam Much Ibnu Subroto, S.T., M.Sc., Ph.D
NIK. 210600017
(Pembimbing)

26-06-2025

Semarang, 12 Juni 2025

Mengetahui,

Kaprodi Teknik Informatika
Universitas Islam Sultan Agung

Moch. Taufik, S.T., MIT
NIK. 210604034

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Ariani Putri Agustina

NIM : 32602100033

Judul Tugas Akhir : Klasifikasi Genre Musik Pop Islami Menggunakan
Convolutional Neural Network (CNN)

Dengan bahwa ini saya menyatakan bahwa judul dan isi Tugas Akhir yang saya buat dalam rangka menyelesaikan Pendidikan Strata Satu (S1) Teknik Informatika tersebut adalah asli dan belum pernah diangkat, ditulis ataupun dipublikasikan oleh siapapun baik keseluruhan maupun sebagian, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka, dan apabila di kemudian hari ternyata terbukti bahwa judul Tugas Akhir tersebut pernah diangkat, ditulis ataupun dipublikasikan, maka saya bersedia dikenakan sanksi akademis. Demikian surat pernyataan ini saya buat dengan sadar dan penuh tanggung jawab.

Semarang, 6 Agustus 2025

Yang Menyatakan,



Ariani Putri Agustina

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Ariani Putri Agustina
NIM : 32602100033
Program Studi : Teknik Informatika
Fakultas : Teknologi industri
Alamat Asal : Pati, Jawa Tengah

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul :
"Klasifikasi Genre Musik Pop Islami Menggunakan Convolutional Neural
Network (CNN)" Menyetujui menjadi hak milik Universitas Islam Sultan Agung
serta memberikan Hak bebas Royalti Non-Eksklusif untuk disimpan,
dialihmediakan, dikelola dan pangkalan data dan dipublikasikan diinternet dan
media lain untuk kepentingan akademis selama tetap menyantumkan nama
penulis sebagai pemilik hak cipta. Pernyataan ini saya buat dengan sungguh-
sungguh. Apabila dikemudian hari terbukti ada pelanggaran Hak
Cipta/Plagiarisme dalam karya ilmiah ini, maka segala bentuk tuntutan hukum
yang timbul akan saya tanggung secara pribadi tanpa melibatkan Universitas
Islam Sultan agung.

Semarang, 6 Agustus 2025

Yang menyatakan,



Ariani Putri Agustina

KATA PENGANTAR

Segala puji dan syukur saya panjatkan ke hadirat Allah SWT atas limpahan rahmat dan karunia-Nya, sehingga saya dapat menyelesaikan Tugas Akhir ini dengan judul “Klasifikasi Genre Musik Pop Islami Menggunakan *Convolutional Neural network* (CNN)”. Tugas Akhir ini disusun sebagai salah satu syarat kelulusan dalam menempuh studi serta untuk memperoleh gelar sarjana (S-1) pada Program Studi Teknik Informatika, Fakultas teknologi Industri, Universitas Islam Sultan Agung Semarang.

Dalam penyusunan Tugas Akhir ini, saya mendapatkan banyak bantuan, baik dalam aspek materi maupun teknis dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, saya ingin mengucapkan terima kasih kepada :

1. Rektor UNISSULA Bapak Prof. Dr. H. Gunarto, SH., M.Hum yang telah memberikan kesempatan penulis menimba ilmu di kampus ini
2. Dekan Fakultas Teknologi Industri Ibu Dr. Ir. Hj. Novi Marlyana, S.T., M.T., IPU., ASEAN Eng
3. Dosen Pembimbing Bapak Imam Much Ibnu Subroto, ST., M.Sc., Ph.D. yang telah meluangkan waktu, membimbing, dan memberi ilmu.
4. Orang tua penulis Bapak Suparmin dan Ibu Ro'aina yang selalu memberikan restu, dukungan, serta Doa dalam menyelesaikan Tugas Akhir ini.
5. Orang terdekat penulis yang senantiasa menemani penulis dalam keadaan sulit dan senang, dan memberikan dukungan kepada penulis.

Penulis menyadari bahwa dalam penyusunan laporan ini masih terdapat banyak kekurangan, untuk itu penulis mengharap kritik dan saran dari pembaca untuk sempurnanya laporan ini. Semoga dengan ditulisnya laporan ini dapat menjadi sumber ilmu bagi setiap pembaca.

Semarang, 27 Mei 2025

Ariani Putri Agustina

DAFTAR ISI

COVER	i
HALAMAN JUDUL	ii
LEMBAR PENGESAHAN TUGAS AKHIR	iii
SURAT PERNYATAAN KEASLIAN TUGAS AKHIR.....	iv
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR GAMBAR.....	ix
DAFTAR TABEL	x
ABSTRAK	xi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan masalah.....	4
1.3 Pembatan masalah.....	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
1.6 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI.....	7
2.1 Tinjauan Pustaka.....	7
2.1 Dasar Teori.....	9
2.2.1 Musik	9
2.2.2 Genre musik	9
2.2.3 Musik Islami	9
2.2.4 Fitur audio untuk klasifikasi.....	10
2.2.5 <i>Mel Frequency Cepstral Coefficients (MFCC)</i>	11
2.2.6 <i>Deep Learning</i>	11
2.2.7 <i>Convolutional Neural network (CNN)</i>	12
2.2.8 <i>Dataset</i>	13
2.2.9 <i>Librosa</i>	14

2.2.10	Evaluasi Model.....	15
BAB III METODE PENELITIAN		17
3.1	Metode penelitian.....	17
3.1.1	Mulai	17
3.1.2	Studi Literatur	17
3.1.3	Pengumpulan Data	18
3.1.4	Ekstraksi Fitur Audio	20
3.1.5	Pembagian Dataset	20
3.1.6	Pemilihan Model	20
3.1.7	Pelatihan Model	20
3.1.8	Implementasi	20
3.1.9	Evaluasi Model.....	20
3.1.10	Selesai	21
BAB IV HASIL DAN PEMBAHASAN.....		22
4.1	Hasil Penelitian	22
4.1.1	Pengumpulan Data	22
4.1.2	Augmentasi Data	23
4.1.3	Ekstraksi Fitur Audio	24
4.1.4	Visualisasi Fitur Audio	26
4.1.5	Persiapan Data dan Pembagian Dataset	28
4.1.6	Persiapan Data untuk Model CNN.....	30
4.1.7	Pemilihan Model	30
4.1.8	Pelatihan Model	32
4.1.9	Evaluasi Model.....	33
4.1.10	Implementasi Model.....	38
BAB V KESIMPULAN DAN SARAN		43
5.1	Kesimpulan	43
5.2	Saran	44
DAFTAR PUSTAKA		
LAMPIRAN		

DAFTAR GAMBAR

Gambar 2. 1 Lapisan <i>Convolutional Neural network</i> (CNN)	12
Gambar 3. 1 flowchart penelitian.....	17
Gambar 3. 2 Visualisasi sinyal audio	19
Gambar 4. 1 folder dataset	22
Gambar 4. 2 Augmentasi Data	24
Gambar 4. 3 ekstraksi fitur audio	25
Gambar 4. 4 visualisasi MFCC	26
Gambar 4. 5 Visualisasi <i>Chroma</i>	27
Gambar 4. 6 Visualisasi <i>Mel-Spectrogram</i>	27
Gambar 4. 7 Potongan kode untuk konversi label dan split dataset.....	29
Gambar 4. 8 Distribusi jumlah data pelatihan, validasi, dan pengujian.....	29
Gambar 4. 9 <i>reshape</i> dan <i>one hot encoding label</i>	30
Gambar 4. 10 output model cnn.....	31
Gambar 4. 11 latih model.....	32
Gambar 4. 12 grafik nilai akurasi.....	34
Gambar 4. 13 grafik nilai loss	35
Gambar 4. 14 <i>confusion matrix</i> hasil prediksi model	37
Gambar 4. 15 penginstalasian <i>ngrok</i>	39
Gambar 4. 16 tampilan awal sistem klasifikasi genre musik Islami dan Pop	39
Gambar 4. 17 tampilan file audio di proses	40
Gambar 4. 18 tampilan hasil prediksi genre islami	41
Gambar 4. 19 Tampilan hasil prediksi genre pop.....	42

DAFTAR TABEL

Tabel 4. 1	Tabel <i>Classification report</i>	36
------------	--	----



ABSTRAK

Genre musik merupakan salah satu elemen penting dalam industri hiburan dan memiliki fungsi untuk menyampaikan pesan, identitas, serta nilai budaya tertentu. Musik Islami sebagai salah satu genre religius memiliki karakteristik khusus, namun sering kali memiliki kemiripan pola dengan musik populer seperti Pop. Hal ini menimbulkan tantangan dalam proses klasifikasi otomatis. Penelitian ini bertujuan untuk membangun sistem klasifikasi genre musik Pop Islami dan Pop menggunakan metode Convolutional Neural Network (CNN) berbasis fitur audio. Data dikumpulkan dari Spotify dan diperluas melalui teknik augmentasi untuk meningkatkan keberagaman data. Selanjutnya, dilakukan ekstraksi fitur menggunakan MFCC, *Chroma*, dan *Mel Spectrogram* guna merepresentasikan ciri khas audio secara numerik. Hasil pelatihan model menunjukkan akurasi sebesar 0.74 (74%). Nilai ini dipengaruhi oleh keterbatasan jumlah data pelatihan, distribusi data yang belum seimbang, serta kemiripan elemen musikal antara kedua genre yang menyebabkan model kesulitan membedakan secara konsisten. Dengan capaian tersebut, sistem yang dibangun memberikan dasar awal yang cukup baik untuk pengembangan ke depan, khususnya melalui peningkatan kualitas data dan penyempurnaan arsitektur model agar hasil klasifikasi menjadi lebih akurat.

Kata kunci: Klasifikasi Musik, CNN, Genre Islami, Genre Pop, Ekstraksi Fitur Audio

ABSTRACT

Music genre is one of the important elements in the entertainment industry and has a role in conveying certain messages, identities, and cultural values. Islamic music as one of the religious genres has specific characteristics, but often has similar patterns with popular music such as Pop. This poses a challenge in the automatic classification process. This research aims to build a classification system for Islamic pop and Pop music genres using the Convolutional Neural Network (CNN) method based on audio features. Data is collected from Spotify and expanded through augmentation techniques to increase data diversity. Furthermore, feature extraction is carried out using MFCC, Chroma, and Mel Spectrogram to numerically represent audio characteristics. The model training results showed an accuracy of 0.74 (74%). This value was influenced by the limited amount of training data, the unbalanced distribution of data, and the similarity of musical elements between the two genres that made it difficult for the model to distinguish consistently. With these achievements, the system provides a good initial basis for future development, especially through improving data quality and refining the model architecture so that classification results become more accurate.

Keywords: Music Classification, CNN, Islamic Genre, Pop Genre, Audio Feature Extraction

BAB I

PENDAHULUAN

1.1 Latar Belakang

Musik merupakan salah satu bentuk seni yang paling umum dan memiliki peran penting dalam kehidupan manusia.. Musik tidak hanya menjadi sarana hiburan, tetapi juga alat untuk menyampaikan pesan, membangun identitas budaya dan mempererat hubungan sosial. Setiap genre musik, seperti pop, jazz, klasik, hingga religi, memiliki ciri khas yang mencerminkan pesan, suasana atau nilai tertentu yang ingin disampaikan oleh penciptanya. Musik telah menjadi bagian tidak terpisahkan dari kehidupan modern dan berkembang pesat berkat kemajuan digital.

Pada saat ini industri musik telah terpengaruh oleh kemajuan digital. Layanan streaming musik menjadi salah satu industri yang sedang berkembang dengan cepat, semakin membantu orang di seluruh dunia mendapatkan akses ke berbagai jenis musik. *Platform streaming* musik yang menyediakan jutaan lagu seperti *Spotify*, *Apple music*, *SoundCloud*, *Youtube Music*, dan sebagainya, yang dapat diakses dengan mudah oleh pengguna di seluruh dunia. Di Tengah perkembangan genre musik yang lain, musik Islami juga menunjukkan eksistensi sebagai salah satu jenis musik yang memiliki penggemarnya tersendiri di tengah perkembangan musik yang lain. Musik Islami seperti, nasyidah, sholawat, qasidah dan lagu religi lainnya, memiliki peran yang strategis dalam menyampaikan pesan moral, nilai-nilai keagamaan, dakwah yang relevan bagi umat muslim.

Selain peran di atas, musik islami juga memiliki kemampuan untuk menciptakan suasana spiritual yang menenangkan serta mempererat hubungan emosional umat muslim. Dengan lirik yang memiliki makna dan melodi yang mendalam, musik Islami dapat dijadikan sarana introspeksi diri dan penguatan iman. Hal ini menjadikan musik Islami tidak hanya relevan untuk kebutuhan hiburan, tetapi juga sebagai alat penting untuk mengajar dan membangun karakter Islami dalam diri.

Sebagai Universitas Islam Sultan Agung (UNISSULA) berkomitmen kuat untuk mengintegrasikan nilai-nilai Islami dalam berbagai aktivitas akademik dan non-akademik. Strategi Pendidikan di Universitas Islam Sultan Agung (Unissula) dirumuskan dengan nama Budaya Akademik Islami atau bisa disebut dengan “BudAI” yang pada intinya berisi penguatan ruhiyyah dan penguatan iptek, dan di Unissula sendiri juga memiliki tema Pendidikan yaitu “Membangun Generasi Khaira Ummah”. Dengan strategi Pendidikan dan tema sentral yang seperti di atas, di Unissula sudah bersahabat baik dengan musik Islami. Musik Islami sering kali digunakan dalam berbagai kegiatan di lingkungan kampus, mulai dari Pengajian, perayaan hari besar Islam, hingga acara mahasiswa. Musik Islami di Unissula tidak hanya berfungsi sebagai hiburan, tetapi juga berfungsi sebagai alat untuk berdakwah dan menyebarkan nilai-nilai moral Islami. Oleh karena itu, sangat penting untuk mengelola dan mengelompokkan musik Islami untuk mendukung keberlangsungan misi Unissula sebagai kampus yang berlandaskan nilai-nilai Islam.

Namun, hingga saat ini musik Islami masih sering dikelompokkan ke dalam genre tertentu secara manual. Proses manual ini tidak hanya memakan waktu, tetapi juga cenderung bersifat subjektif. Selain itu, penelitian tentang pengelompokan musik Islami berbasis teknologi masih jarang dilakukan. Dengan adanya sistem otomatis, Unissula dapat membuka peluang besar dan menciptakan database musik Islami yang lebih terorganisir untuk mendukung dakwah Islami melalui musik. Dengan berkembangnya teknologi kecerdasan buatan, muncul peluang untuk mengotomisasi proses ini dengan lebih akurat dan efisien. Salah satu pendekatan teknologi yang relevan adalah menggunakan Convolutional Neural Networks (CNN). CNN telah banyak digunakan dan dikenal karena memiliki keunggulan dalam pengenalan pola data visual dan telah banyak diterapkan untuk klasifikasi gambar maupun audio.

Dalam beberapa penelitian sebelumnya, teknologi CNN telah menunjukkan hasil yang sangat baik dalam klasifikasi genre musik. (K. Choi, 2017) dalam penelitiannya yang berjudul “*Convolutional Recurrent Neural Networks for Music Classification*”, menunjukkan bahwa pendekatan *Deep Learning* berbasis CNN

mampu meningkatkan akurasi klasifikasi genre musik secara signifikan dibandingkan dengan metode tradisional.

Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi genre musik islami berbasis CNN. Sistem ini diharapkan dapat membantu kampus dalam mengelompokkan musik islami secara otomatis dan akurat, sehingga mendukung berbagai kegiatan yang ada di kampus yang berbasis nilai-nilai Islam. Selain itu, hasil penelitian ini juga dapat memberikan kontribusi antara lain dalam pengembangan teknologi pengolahan musik Islam di era digital.

Dalam konteks penelitian ini, musik islami didefinisikan sebagai lagu-lagu yang memiliki nuansa atau karakteristik khas musik Islami, tanpa mempertimbangkan isi lirik secara eksplisit. Artinya, klasifikasi dilakukan berdasarkan ciri-ciri audio seperti melodi, harmoni, dan struktur musik, bukan dari makna kata-kata atau teks yang dinyanyikan.

Penelitian ini tidak menerapkan ekstraksi lirik, melainkan fokus pada pola akustik dan fitur audio yang dapat membedakan genre Islami dan Pop. Fitur-fitur yang digunakan, seperti MFCC, Chroma, dan Mel Spectrogram, mewakili bentuk representasi numerik dari karakter suara, yang kemudian digunakan untuk melatih model klasifikasi.

Pendekatan ini diambil karena lagu Islami umumnya memiliki pola musikal tertentu, misalnya tempo yang lebih tenang, instrumen khas seperti gambus atau rebana, serta struktur musik yang cenderung konsisten. Dengan demikian, meskipun tanpa melihat isi lirik, model tetap dapat mempelajari pola audio yang umum dijumpai pada lagu-lagu Islami dan menggunakannya sebagai dasar klasifikasi.

Dengan demikian, penelitian ini tidak hanya relevan untuk mendukung kegiatan akademik di Unissula, tetapi juga akan berdampak luas pada pengembangan musik Islami di era teknologi modern. Implementasi teknologi CNN akan meningkatkan efisiensi dan objektivitas dalam pengelolaan musik Islami, mendukung kegiatan dakwah, dan mencerminkan bagaimana nilai-nilai Islam diintegrasikan dengan kemajuan teknologi yang menjadi identitas Unissula.

1.2 Perumusan masalah

Berdasarkan uraian latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana Teknik *Deep Learning* dapat digunakan untuk mengembangkan model klasifikasi otomatis untuk musik islami
2. Bagaimana kinerja model yang dikembangkan untuk klasifikasi musik islami berdasarkan data audio yang tersedia

1.3 Pembatasan masalah

Adapun Batasan masalah penelitian ini adalah:

1. Penelitian menggunakan data audio dalam format tertentu, seperti MP3 atau WAV. Penelitian akan fokus pada lagu dengan durasi pendek, tidak mempertimbangkan cuplikan musik yang pendek
2. Mempertimbangkan hanya menggunakan genre genre populer seperti, pop
3. Dalam penelitian ini, data yang digunakan merupakan hasil dari pengumpulan dan penyusunan data secara mandiri oleh peneliti.
4. Penelitian berfokus pada ekstraksi fitur audio standar seperti *Mel-Frequency Cepstral Coefficients* (MFCC), *chroma*, *mel-spectrogram* atau beberapa fitur lain yang sering digunakan dalam klasifikasi audio. Pemanfaatan fitur yang lebih canggih akan menjadi pertimbangan dalam pengembangan di masa depan.
5. Penelitian tidak mempertimbangkan pada lirik lagu hanya memfokuskan audio

1.4 Tujuan Penelitian

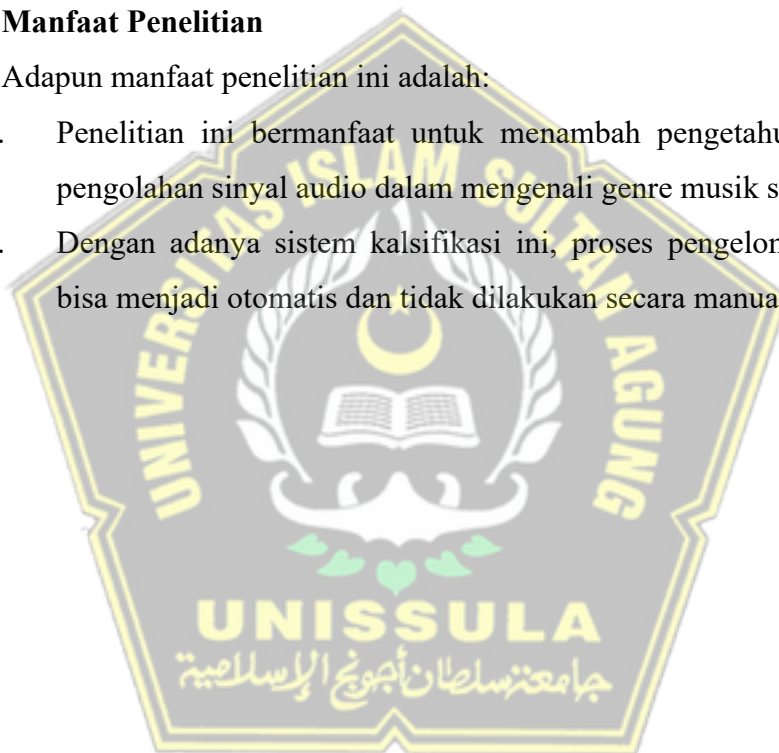
Adapun tujuan penelitian ini adalah:

1. Merancang model klasifikasi yang dapat mengidentifikasi musik islami secara otomatis
2. Mengevaluasi kinerja model CNN dengan metrik, seperti akurasi presisi, *recall*, dan *F1-score*, untuk memastikan bahwa model tersebut memiliki kemampuan klasifikasi terbaik

1.5 Manfaat Penelitian

Adapun manfaat penelitian ini adalah:

1. Penelitian ini bermanfaat untuk menambah pengetahuan di bidang pengolahan sinyal audio dalam mengenali genre musik secara otomatis
2. Dengan adanya sistem klasifikasi ini, proses pengelompokan musik bisa menjadi otomatis dan tidak dilakukan secara manual



1.6 Sistematika Penulisan

Sistematika penulisan yang akan digunakan oleh penulis dalam sebuah pembuatan laporan tugas akhir ini adalah sebagai berikut:

BAB I : PENDAHULUAN

Pada Bab 1, penulis mengutarakan urgensi dari penelitian yang diangkat, mulai dari penulisan latar belakang, membuat rumusan masalah, membatasi permasalahan yang dibahas, serta tujuan dan manfaat yang diperoleh, dan sistematika penulisan.

BAB II : TINJAUAN PUSTAKA DAN DASAR TEORI

Pada bab 2, penulis membuat dasar teori yang digunakan, serta rujukan dari penelitian terdahulu yang akan digunakan dalam perancangan sistem, dan membantu penulis untuk memahami teori yang berhubungan dengan metode Algoritma *Convolutional Neural Network* selama proses penelitian.

BAB III : METODE PENELITIAN

Pada bab 3, penulis mengungkapkan proses dan tahapan penelitian yang dimulai dari mendapatkan dataset hingga proses pemodelan topik data yang ada.

BAB IV : HASIL DAN ANALISIS PENELITIAN

Pada bab 4, penulis mengungkapkan hasil penelitian, dimulai dengan hasil akhir sistem, klasifikasi data uji, dan akurasi dari sistem.

BAB V : KESIMPULAN DAN SARAN

Pada bab 5, penulis memaparkan kesimpulan dari hasil proses penelitian mulai dari awal sampai akhir.

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Hasil penelitian terdahulu memiliki peran penting sebagai dasar pendukung dalam pelaksanaan penelitian selanjutnya. Di satu pihak, kajian tersebut dapat memberikan referensi pembanding yang bermanfaat terhadap studi-studi sebelumnya, serta membantu dalam mengenali keunggulan dan kelemahannya. Selain itu, telaah terhadap penelitian terdahulu juga berfungsi untuk memperkuat argumen yang disusun dalam penelitian baru. Hal ini memungkinkan peneliti memperoleh pemahaman yang lebih mendalam mengenai isu-isu yang telah diteliti sebelumnya dalam konteks yang serupa. Berikut ini merupakan beberapa penelitian yang relevan:

Adapun penelitian sebelumnya (Ma, 2023) melakukan penelitian klasifikasi genre musik melalui berbagai Teknik Machine Learning termasuk *K-Nearest Neighbors*, *Random Forest Classifier* (RFC), *Artificial Neural Networks* (ANN) dan *Convolutional Neural Networks* (CNN). Menggunakan dataset dari GTZAN dengan 1000 sampel audio dari 10 genre, penelitian ini menekankan akan pentingnya ekstraksi fitur dan pemilihan model dalam meningkatkan akurasi klaisifikasi. Model KNN mencapai kaurasi sebesar 43%, model ANN mencapai sekitar 69%, dan yang paling unggul adalah model CNN dengan akurasi 80%. Penelitian ini juga membahas tantangan seperti overfitting dan menyarankan agar menggabungkan CNN dengan recurrent neural networks (RNN) dapat meningkatkan kinerja.

Kemudian dari hasil penelitian (Guo, 2024) melakukan pengenalan genre musik melalui *Deep Learning* yang menekankan keuntungan dari ekstrasi fitur otomatis dan penimngkatan akurasu dibandingkan metode tradisional. Penelitian ini menggunakan *Convolutional Neural Networks* (CNNs) dan *Recurrent Neural Netwroks* (RNNs) untuk kalsifikasi genre yang efektif, memununjukkan peningkatan siginifikan dalam akurasi dan efisiensi. Studi ini juga mmebahasa tantangan dalam mengintegrasikan persepsi manusia yang subjektif dan

memahami fitur diskriminatif, sambil menyoroti pentingnya metode ekstraksi dan penyetelan model. Arah penelitian di masa depan mencakup pengoptimalan model dan peningkatan interpretabilitas dalam aplikasi pengenalan genre musik.

Kemudian dari penelitian (He & Li, 2024) yang membahas tentang kemajuan dalam Teknik pemrosesan audio, khususnya yang berfokus pada klasifikasi musik dan metode overlay menggunakan pendekatan *Deep Learning*. Ini menyoroti pentingnya fitur audio seperti *over-zero* dan *Mel-frequency cepstrum coefficients* dalam membedakan suara, mengkritik metode klasifikasi musik tradisional karena keterbatasannya dalam menangkap fitur musik yang kompleks dan mengusulkan model baru yang menggabungkan antara CNN dan RNN untuk meningkatkan akurasi klasifikasi. Studi ini menekankan pentingnya data quality, preprocessing steps, dan integrasi berbagai model *Deep Learning* untuk meningkatkan ketahanan dan efektivitas klasifikasi musik.

Kemudian penelitian (Kumar dkk., 2023) yang menyajikan metode untuk klasifikasi genre musik secara otomatis dengan memanfaatkan *Convolutional Neural Networks* (CNN) dan *Mel-Frequency Cepstral Coefficients* (MFCC) untuk ekstraksi fitur dari sinyal audio. Studi ini menggunakan dataset GTZAN, yang terdiri dari 1.000 klip audio di sepuluh genre, dan melaporkan akurasi sebesar 83%, melebihi beberapa metode tradisional. Dokumen ini membahas metrik evaluasi yang digunakan termasuk akurasi, *F1-Score*, *presisi*, *recall*, dan *confusion matrix*, serta menyarankan peningkatan di masa depan seperti penanganan panjang audio yang bervariasi, *supporting multiple formats*, augmentasi data, hyperparameter tuning untuk lebih meningkatkan kinerja model.

Kemudian dari hasil penelitian (Fardhani dkk., 2021) yang membahas klasifikasi genre musik menggunakan *Convolutional Neural network* (CNN) dengan dua metode ekstraksi fitur yaitu *Mel-Frequency Cepstral Coefficient* (MFCC) dan *spectrogram*. Penelitian ini menganalisis metrik kinerja berbagai genre musik, mengungkapkan bahwa Jazz memiliki kinerja terbaik secara keseluruhan, sementara *Country* memiliki metrik terendah. Selain itu, penelitian ini menekankan penggunaan *spectrogram* dengan mencapai akurasi 76% dalam klasifikasi genre musik, dan menekankan kebutuhan akan metode yang lebih baik

dalam *Music Information Retrieval* (MIR) dengan menyarankan penambahan data yang lebih banyak dan penggabungan metode ekstraksi fitur yang lebih detail.

2.1 Dasar Teori

2.2.1 Musik

Musik merupakan salah satu bentuk hiburan yang digemari oleh banyak orang. Umumnya, musik dimanfaatkan sebagai media untuk menyalurkan perasaan atau emosi manusia (Brawijaya dkk., 2017). Pengertian musik sangat beragam, di antaranya menyatakan bahwa musik adalah bunyi atau kesan yang diterima oleh indra pendengaran. Selain itu, musik juga dipandang sebagai karya seni yang melibatkan unsur utama dan pendukung dari berbagai usia, serta sebagai bunyi yang diciptakan secara sengaja oleh individu maupun kelompok yang dipersembahkan dalam bentuk musik (Wairata dkk., 2021).

2.2.2 Genre musik

Genre musik merupakan sebuah istilah yang dibuat serta dimanfaatkan oleh manusia sebagai cara untuk mengelompokkan dan menggambarkan jenis-jenis musik. Tidak terdapat batasan yang kaku dalam mendefinisikan genre musik, sebab genre terbentuk dari berbagai interaksi yang saling memengaruhi antara aspek sosial, pemasaran, sejarah, hingga budaya. Jenis-jenis genre musik sendiri sangat beragam, di antaranya yaitu musik klasik, jazz, blues, country, techno, reggae, R&B, dangdut, pop, dan lain (Ayu dkk., 2023).

2.2.3 Musik Islami

Musik Islami merujuk pada jenis musik yang memiliki lirik dan tema yang terkait dengan Islam, seperti pengagungan Tuhan, rasul, atau pesan-pesan moral dan spiritual. Musik Islami bisa bervariasi dari genre yang berbeda, termasuk nasyid (vokal tanpa instrumen musik), musik instrumen tradisional, atau bahkan modern seperti hip-hop atau pop dengan lirik yang Islami. Penting untuk dicatat bahwa pendapat tentang musik dalam Islam dapat berbeda-beda, tergantung pada

interpretasi dan pemahaman masing-masing individu atau kelompok.(Kurniawan Soeganda & Budiyan, 2024).

2.2.4 Fitur audio untuk klasifikasi

Fitur musik yang berbasis konten diperoleh melalui ekstraksi dari sinyal audio. Beragam fitur tersebut dapat dinyatakan dalam bentuk nilai rata-rata (*mean*), deviasi standar, *variance*, histogram, agregasi MFCC, serta *area moments*. Secara garis besar, fitur musik berbasis konten dibedakan menjadi dua kategori, yaitu fitur domain waktu dan fitur domain frekuensi.(Ayu dkk., 2023)

Fitur pada domain waktu diperoleh secara langsung dari sinyal audio. Tiga fitur utama dalam kategori ini mencakup :

- a. *Root mean square energy*, yakni nilai rata-rata akar kuadrat dari energi dalam sinyal audio diskrit.
- b. *Zero crossing rate*: jumlah perubahan tanda sinyal dari positif ke negatif (atau sebaliknya) dalam sinyal waktu diskrit.
- c. *Tempo* yang menggambarkan kecepatan musik dan biasanya dinyatakan dalam satuan ketukan per menit.

Sementara itu, fitur domain frekuensi diperoleh dengan menerapkan transformasi Fourier pada sinyal audio. Beberapa fitur yang termasuk dalam kelompok ini antara lain

- a. *Mel-Frequency Cepstral Coefficient* (MFCC): menggambarkan spektrum energi jangka pendek dalam file audio dan mencerminkan karakteristik suara manusia.
- b. *Chroma*: berkaitan dengan distribusi energi sinyal pada 12 nada dalam satu oktaf (*semitone*).
- c. *Spectral centroid*: menunjukkan titik tengah energi spektrum pada suatu frame audio.
- d. *Spectral bandwidth*: menggambarkan rentang frekuensi dari suatu sinyal bunyi.
- e. *Spectral Roll-off*: frekuensi di mana sejumlah tertentu (biasanya dalam persen) dari total energi spektrum terkonsentrasi di bawahnya.

2.2.5 *Mel Frequency Cepstral Coefficients (MFCC)*

MFCC merupakan representasi dari *mel-spectrogram*. MFCC menampilkan fitur-fitur yang mendeskripsikan bentuk frekuensi dari spektrum audio. Langkah-langkah proses MFCC berdasarkan pada perbedaan frekuensi yang didengar oleh indra pendengaran manusia, sehingga fitur ini seperti manusia merepresentasikan sinyal suara. MFCC berfungsi untuk mengubah sinyal audio menjadi array multidimensi (Pratama, 2021).

Spektrogram mampu merepresentasikan berbagai frekuensi dalam sinyal suara. Namun demikian, pendekatan ini mengikuti pola kerja koklea dalam sistem pendengaran manusia, yang menangkap frekuensi pada titik-titik tertentu. Untuk menyesuaikan dengan karakteristik tersebut, dikembangkan suatu formulasi yang menyerupai respons pendengaran manusia, dikenal dengan sebutan skala mel (*mel-scale*).

Skala mel bekerja dengan menerapkan penyaringan (*filtering*) terhadap koefisien frekuensi yang dihasilkan dari spektrogram. Proses penyaringan ini bersifat linier hingga frekuensi 1000 Hz, dan berubah menjadi logaritmik untuk frekuensi di atasnya. Metode inilah yang digunakan dalam perhitungan mel-frekuensi (Wairata dkk., 2021).

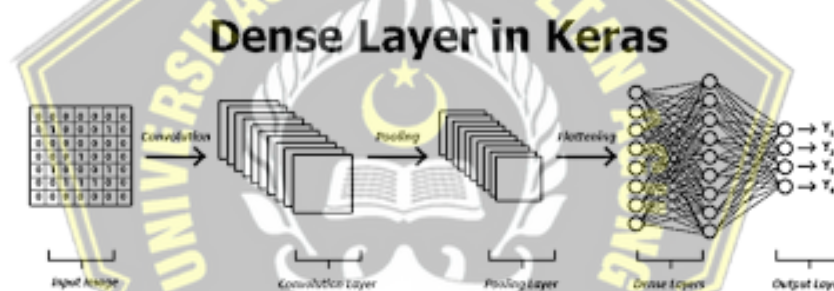
2.2.6 *Deep Learning*

Pembelajaran dalam (bahasa Inggris: *Deep Learning*) atau sering dikenal dengan istilah pembelajaran struktural mendalam (bahasa Inggris: *deep structured learning*) atau pembelajaran hierarki (bahasa Inggris: *hierarchical learning*) adalah salah satu cabang dari ilmu pembelajaran mesin (bahasa Inggris: *Machine Learning*) yang terdiri algoritme pemodelan abstraksi tingkat tinggi pada data menggunakan sekumpulan fungsi transformasi *non-linear* yang ditata berlapis-lapis dan mendalam. *Deep Learning* merupakan cabang ilmu *Machine Learning* berbasis Jaringan Saraf Tiruan (JST) atau bisa dikatakan sebagai perkembangan dari JST. Dalam *Deep Learning*, sebuah komputer belajar mengklasifikasi secara langsung dari gambar atau suara. *Convolutional Neural network* (CNN/ConvNet) adalah salah satu algoritma *Deep Learning* yang merupakan pengembangan dari *Multilayer Perceptron* (MLP) yang dirancang untuk mengolah data dalam bentuk

dua dimensi, misalnya gambar atau suara. CNN dapat belajar langsung dari citra sehingga mengurangi beban dari pemrograman. Metode *Deep Learning* menggunakan *CPU* dan *RAM* dalam proses komputasi, dan juga memanfaatkan *GPU* sehingga proses komputasi data yang besar dapat berlangsung lebih cepat (Wairata dkk., 2021).

2.2.7 Convolutional Neural network (CNN)

Convolutional Neural network (CNN) merupakan salah satu jenis neural network yang biasa digunakan pada *image classification*. Metode kecerdasan buatan ini terinspirasi dari pendistribusian neuron yang ada pada otak biologis. CNN adalah turunan dari *Multilayer Perceptron* (MLP) yang dioptimalkan untuk pengenalan pola dua dimensi seperti gambar 2.1. CNN menggunakan layer khusus yaitu *Convolution Layer* dan *Pooling Layer* (Pratama, 2021)



Gambar 2. 1 Lapisan *Convolutional Neural network* (CNN)

a. *Convolutional Layer*

Convolution layer digunakan untuk mengekstrak fitur-fitur yang terdapat pada data yang dimasukkan. Pada layer ini proses convolution terjadi, dimana sebuah filter bergerak melintasi seluruh bidang gambar masukan untuk mengambil fitur yang ada pada gambar.

b. *Pooling Layer*

Fungsi penggunaan dari pooling layer adalah mengurangi dimensi (*down sampling*) dari feature map atau keluaran dari convolution layer sehingga dapat mempercepat proses komputasi dan mengurangi *over-fitting*. *Pooling layer* biasanya berada setelah convolution layer. Pooling layer yang biasa digunakan adalah *Max Pooling*, dimana keluaran *array* dari filter merupakan nilai tertinggi

yang dideteksi dan *Average Pooling*, dimana keluaran merupakan hasil rata-rata penjumlahan dari filter.

c. *Batch Normalization*

Batch normalization merupakan teknik normalisasi yang terjadi diantara layer. Normalisasi ini berfungsi untuk mengubah nilai numerik yang masuk ke skala tertentu sebelum meneruskannya ke layer berikutnya.

d. *Fully Connected Layer*

Fully-Connected (FC) layer berfungsi mengklasifikasi berdasarkan fitur yang diekstraksi dari layer sebelumnya. Pada lapisan ini setiap neuron terhubung dengan neuron yang berada pada lapisan sebelum dan selanjutnya, sesuai dengan namanya. Karena feature map yang dihasilkan layer sebelumnya masih berupa *array* multidimensi, maka harus diubah menjadi array satu dimensi sebelum bisa digunakan sebagai masukan dari *fully connected layer*.

2.2.8 Dataset

Dataset merupakan kumpulan data yang disimpan di dalam memori (in-memory) dan berfungsi layaknya sebuah basis data. Dataset juga dapat dimaknai sebagai kumpulan objek yang merepresentasikan data beserta relasinya di dalam memori. Secara struktur, dataset memiliki kemiripan dengan data yang disimpan di dalam database. Dataset ini terdiri dari sekumpulan tabel serta data di dalamnya (Yahya & Mahpuz, 2019).

Dataset ada dua macam yaitu :

1. *Private Dataset* : jenis dataset yang diperoleh dari institusi atau organisasi tertentu yang menjadi lokasi atau objek penelitian. Contoh dari dataset ini misalnya berasal dari rumah sakit, instansi pemerintah, pabrik, atau perusahaan jasa (Suhartini dkk., 2020).
2. *Public Dataset* : jenis dataset yang tersedia secara umum dan dapat diakses melalui repository publik yang telah disepakati penggunaannya oleh para peneliti. Salah satu contohnya adalah dataset dari UCI (Suhartini dkk., 2020).

2.2.9 *Librosa*

Librosa adalah Pustaka python yang dirancang khusus untuk analisis dan pemrosesan sinyal audio dan musik. *Librosa* adalah alat ekstraksi fitur lain dengan kemampuan ekstraksi fitur secara real-time. Langkah pertama menuju analisis fitur audio dan ekstraksi fitur adalah memuat file audio ke dalam kode. Hal ini dilakukan dengan menggunakan fungsi `load` yang mengambil dua argument utama sebagai nama file audio dan nilai *sr* (*sampling rate*). File audio yang dimuat dan disimpan dalam format array dapat digunakan untuk memvisualisasikan file audio menggunakan `plot` (Singh dkk., 2021).

Menurut (McFee dkk., 2015), *Librosa* bertujuan untuk menyediakan antarmuka yang konsisten dan mudah digunakan untuk analisis sinyal audio, yang sangat berguna dalam bidang *Music Information Retrieval* (MIR), pengenalan ucapan dan klasifikasi genre musik.

Librosa sendiri mempunyai beberapa fitur utama (Khamitkar & Lokhande, 2021):

- a. Ekstraksi fitur audio : *Librosa* memiliki ekstraksi berbagai fitur audio seperti *Mel-Frequency Cepstral Coefficients* (MFCC), *chroma features*, *spectral contrast*, dan *zero-crossing rate*, yang sering digunakan dalam tugas-tugas klasifikasi dan analisis audio.
- b. Visualisasi : *Librosa* juga menyediakan fungsi untuk memvisualisasikan sinyal audio dalam domain waktu dan frekuensi, seperti *waveform*, *spektrogram* dan *diagram kromagram*, yang membantu dalam pemahaman karakteristik sinyal audio.
- c. Analisis Ritme dan Tempo : *Librosa* dapat digunakan untuk mendeteksi tempo dan ritme dalam sinyal audio, yang penting dalam analisis musik dan pengenalan pola.
- d. Transformasi Waktu-Frekuensi: Pustaka ini mendukung transformasi seperti *Short-Time Fourier Transform* (STFT) dan *Constant-Q Transform* (CQT), yang digunakan untuk analisis spektral sinyal audio.

Dalam penelitian oleh (McFee dkk., 2015) , *Librosa* digunakan untuk memfasilitasi analisis sinyal musik dan audio dalam *Python*, dengan fokus pada kemudahan penggunaan dan integrasi dengan pustaka *Python* lainnya.

2.2.10 Evaluasi Model

Evaluasi model dilakukan untuk menilai seberapa baik performa sistem klasifikasi dalam memproses data yang belum pernah digunakan sebelumnya oleh model tersebut..

Beberapa metrik yang digunakan untuk evaluasi antara lain:

1. *Accuracy* : presentase prediksi benar terhadap total data. Nilai akurasi dihitung berdasarkan rasio antara jumlah prediksi yang benar terhadap total prediksi yang dilakukan. Persamaan yang digunakan untuk menghitung nilai akurasi (Ayu dkk., 2023).

$$Accuracy = \frac{\text{Number of correct prediction}}{\text{Total of number prediction}} \quad (1)$$

2. *Precision* : ketepatan prediksi model dalam menentukan genre tertentu. Metrik ini menunjukkan seberapa besar bagian dari prediksi positif yang benar-benar sesuai dengan kelas positif yang sebenarnya. Nilai precision diperoleh dari rasio antara jumlah *true positive* terhadap seluruh data yang diklasifikasikan sebagai positif.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

3. *Recall* : kemampuan model menangkap semua data dari genre tertentu. Recall digunakan untuk mengukur sejauh mana model mampu mendeteksi seluruh data yang termasuk dalam kelas positif. Nilai ini dihitung berdasarkan proporsi antara jumlah *true positive* dengan total data yang sebenarnya merupakan kelas positif.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

4. *F1- score* : harmoni antara *precision* dan *recall*. Nilai *F1-Score* yang tinggi merepresentasikan nilai *precision* dan *recall* yang tinggi juga.

$$F1\ Score = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (4)$$

5. *Confusion matrix* : metode evaluasi yang digunakan dalam sistem klasifikasi untuk mengukur akurasi prediksi model berdasarkan jumlah data yang terklasifikasi dengan benar maupun salah. Dalam konteks ini, genre musik dipahami sebagai pengelompokan yang dibentuk dan digunakan manusia untuk menggambarkan serta mengidentifikasi musik. Genre sendiri tidak memiliki batasan atau definisi tunggal, karena lahir dari interaksi kompleks berbagai elemen seperti masyarakat, strategi pemasaran, sejarah, dan nilai-nilai budaya. Konsep *confusion matrix* menyajikan empat kategori utama yang digunakan untuk menghitung hasil klasifikasi, yaitu:

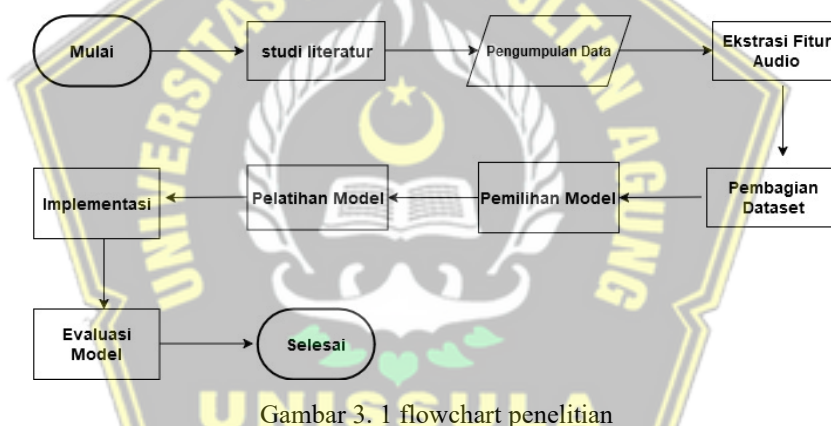
- a. *True positive* (TP) adalah jumlah data dengan label positif yang berhasil dikenali dengan benar sebagai data positif oleh model klasifikasi.
- b. *False positive* (FP) adalah jumlah data berlabel negatif yang secara keliru terklasifikasi sebagai positif
- c. *False negative* (FN) adalah jumlah data berlabel positif yang salah diprediksi sebagai data negatif.
- d. *True negative* (TN) adalah jumlah data berlabel negatif yang diklasifikasikan dengan tepat sebagai negatif.

BAB III

METODE PENELITIAN

3.1 Metode penelitian

Penelitian ini menggunakan pendekatan berbasis *Deep Learning* dengan memanfaatkan arsitektur *Convolutional Neural Network* (CNN) untuk mengklasifikasikan genre musik berdasarkan fitur audio. Proses penelitian ini dilakukan secara bertahap, dimulai dari studi literatur hingga tahap implementasi dan evaluasi model. Setiap langkah dirancang untuk memastikan data audio diolah secara optimal sehingga model mampu mengenali pola yang relevan untuk prediksi genre. Adapun tahapan penelitian yang dilakukan dapat dilihat pada diagram alur berikut.



Gambar 3. 1 flowchart penelitian

3.1.1 Mulai

Penelitian dimulai dengan menetapkan tujuan, rumusan masalah, dan ruang lingkup penelitian. Ini adalah langkah awal untuk memulai proses penelitian dengan arah yang jelas.

3.1.2 Studi Literatur

Dilakukan kajian terhadap literatur dan penelitian sebelumnya yang relevan, untuk memahami konsep, metode, dan algoritma yang telah digunakan, serta untuk mengidentifikasi celah penelitian yang dapat diisi melalui penelitian ini.

3.1.3 Pengumpulan Data

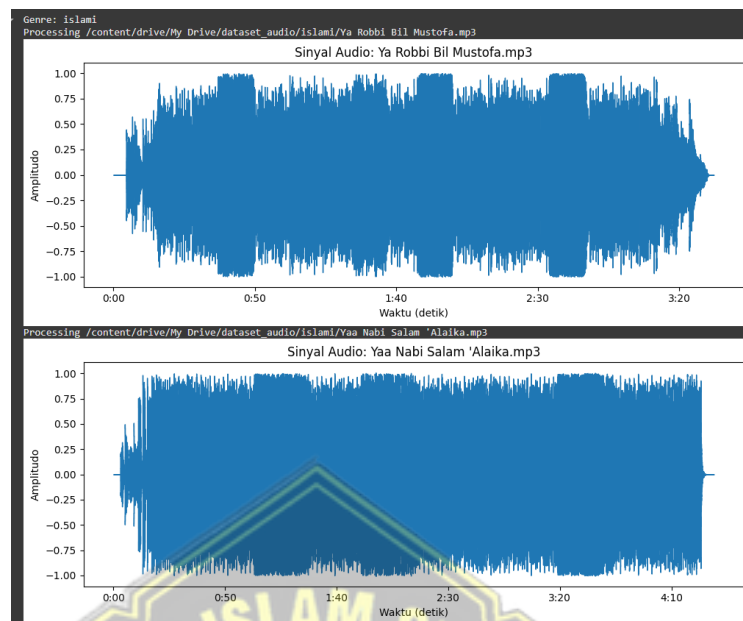
Pada tahap ini, data berupa musik Islami dikumpulkan. Sumber data dapat berupa dataset publik atau dataset yang dibuat sendiri. Misalnya, lagu-lagu Islami diunduh dari platform seperti *Spotify* dan disusun dalam folder terstruktur sesuai dengan genre untuk mempermudah pengolahan data.

Dalam proses pengambilan data untuk penelitian ini, dataset lagu dibuat secara mandiri dengan cara mengunduh lagu-lagu Islami dari platform *Spotify*. Proses ini dilakukan untuk memastikan bahwa dataset yang digunakan sesuai dengan kebutuhan penelitian dan fokus pada genre musik Islami. Setelah lagu-lagu diunduh, file audio tersebut disimpan dalam format MP3 agar kompatibel dengan langkah-langkah pemrosesan audio berikutnya.

File audio yang telah diunduh kemudian dikelompokkan berdasarkan genre atau subkategori tertentu dan disusun ke dalam folder di *Google Drive*. Penggunaan *Google Drive* dipilih karena kemudahannya aksesnya dalam integrasi dengan *Google Colab*. Struktur folder diatur dengan sistem hierarki, di mana folder utama dinamakan sebagai **dataset_audio2**, dan di dalamnya terdapat subfolder berdasarkan nama genre, misalnya pop islami, pop. Setiap subfolder berisi file-file audio yang sesuai dengan kategori tersebut.

Dengan metode ini, peneliti dapat memastikan bahwa data terorganisasi dengan baik sehingga memudahkan proses pengolahan data selanjutnya, seperti membaca file audio, mengekstraksi fitur, dan melakukan pembagian dataset untuk pelatihan dan pengujian model. Struktur yang sistematis juga mengurangi risiko kesalahan dalam pengelolaan data, serta mempermudah replikasi penelitian di masa mendatang.

Pembuatan *dataset* mandiri memungkinkan untuk memastikan bahwa data sesuai dengan kebutuhan spesifik dalam penelitian, termasuk genre, bahasa, atau karakteristik unik yang tidak dapat ditemukan dalam dataset publik. *Dataset* tersebut juga akan digunakan untuk melatih dan menguji model.



Gambar 3. 2 Visualisasi sinyal audio

Gambar 3.2 menampilkan visualisasi sinyal audio dari dua lagu dengan genre musik Islami. Visualisasi ini berupa *waveform*, yang merepresentasikan perubahan amplitudo terhadap waktu untuk setiap file audio. *Waveform* memberikan gambaran tentang karakteristik dasar audio, seperti intensitas suara, dinamika volume, dan durasi lagu. Pada sinyal pertama, durasi lagu sekitar 3 menit 20 detik, sedangkan pada sinyal kedua durasinya lebih panjang, yaitu sekitar 4 menit 10 detik. Kedua sinyal menunjukkan pola amplitudo yang berbeda, mencerminkan variasi dalam dinamika vokal, instrumental, atau struktur musiknya. Amplitudo yang lebih tinggi menunjukkan bagian dengan intensitas suara yang lebih kuat, seperti pada chorus atau klimaks lagu. Visualisasi ini menjadi langkah awal dalam memahami pola gelombang audio, yang dapat digunakan untuk analisis lebih lanjut, seperti ekstraksi fitur audio dan klasifikasi genre musik menggunakan metode berbasis *Deep Learning* seperti *Convolutional Neural network* (CNN).

3.1.4 Ekstraksi Fitur Audio

Data musik yang dikumpulkan kemudian diekstraksi untuk mendapatkan fitur-fitur penting seperti *MFCC (Mel-Frequency Cepstral Coefficients)*, *chroma features*, *mel-spectrogram* dan lain-lain. Fitur ini merepresentasikan karakteristik unik dari setiap lagu secara numerik, yang akan menjadi input bagi model CNN.

3.1.5 Pembagian Dataset

Dataset dibagi menjadi tiga subset utama:

- a. *Training Set*: Untuk melatih model CNN.
- b. *Validation Set*: Untuk mengoptimalkan model selama pelatihan.
- c. *Test Set*: Untuk mengevaluasi kinerja model setelah selesai dilatih.

Pembagian *Dataset* dilakukan untuk melatih model pada Sebagian besar data dan kemudian untuk mengevaluasi performa pada data yang belum pernah dilihat oleh model. Pembagian dataset dibagi menjadi 3 yaitu Data latih sebanyak 70%, data validasi sebanyak 15% dan data uji sebanyak 15%. Dapat menggunakan *Cross-validation* untuk memvalidasi hasil model yang lebih lanjut.

3.1.6 Pemilihan Model

Model CNN dipilih sebagai algoritma klasifikasi utama. CNN dipilih karena kemampuannya yang unggul dalam mengenali pola dari data kompleks seperti fitur audio. Pemilihan ini juga melibatkan desain arsitektur CNN yang sesuai untuk data fitur audio.

3.1.7 Pelatihan Model

Model CNN dilatih menggunakan training set, dengan berbagai parameter seperti *learning rate*, kernel size, dan jumlah filter. Proses ini dilakukan melalui beberapa iterasi (*epoch*) hingga model mencapai akurasi yang optimal.

3.1.8 Implementasi

Model yang telah dilatih diimplementasikan untuk melakukan klasifikasi genre musik Islami. Implementasi melibatkan integrasi *pipeline* dari data audio mentah hingga menghasilkan prediksi genre.

3.1.9 Evaluasi Model

Model diuji menggunakan data test set. Evaluasi dilakukan menggunakan metrik seperti:

- a. Akurasi: Untuk mengukur keakuratan prediksi model.
- b. *Precision, Recall, F1-Score*: Untuk menganalisis kualitas prediksi.
- c. *Confusion Matrix*: Untuk memahami kesalahan dan kinerja model secara detail.

3.1.10 Selesai

Setelah semua tahap selesai, penelitian dinyatakan selesai dengan menyusun laporan dan menyimpulkan hasil.



BAB IV

HASIL DAN PEMBAHASAN

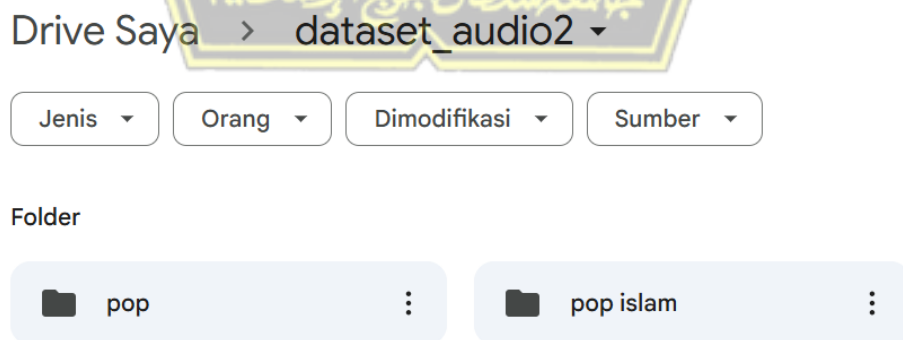
4.1 Hasil Penelitian

Pada bagian ini, peneliti akan membahas mengenai yang didapatkan setelah melalui tahap metode penelitian sebagai berikut :

4.1.1 Pengumpulan Data

Pada tahap ini, peneliti menggunakan *Google Drive* sebagai alat media utama untuk mengatur dan menyusun file-file audio yang akan digunakan dalam penelitian. Data yang dikumpulkan terdiri dari musik yang terbagi dalam dua genre, yaitu musik Islami dan Pop. Proses pengumpulan data dilakukan dengan cara mendengarkan musik tersebut melalui *Spotify* versi web, dengan pemahaman bahwa platform tersebut menyediakan berbagai jenis musik dengan kualitas audio yang baik dan sesuai dengan kriteria yang telah ditetapkan.

Setelah proses selesai, file audio tersebut disimpan di perangkat lokal. Setelah itu, semua file tersebut disimpan secara manual di *Google Drive* untuk memudahkan akses dan integrasi dengan platform penyimpanan data, *Google Colab*. Pemilihan *Google Drive* sebagai tempat penyimpanan didasarkan pada kemudahannya dalam integrasi langsung dengan *Google Colab*, sehingga file dapat diakses tanpa perlu proses unggah ulang secara manual.



Gambar 4. 1 folder dataset

Gambar 4.1 menunjukkan struktur folder dari dataset audio yang digunakan dalam penelitian ini. Dataset disimpan di *Google Drive* dengan nama folder utama **dataset_audio2**, yang berisi dua subfolder yaitu **Pop Islami** dan **Pop**. Masing-masing folder mewakili satu kelas atau genre musik yang menjadi objek klasifikasi dalam penelitian ini. Melalui pendekatan ini, proses analisis fitur audio seperti *MFCC (Mel Frequency Cepstral Coefficient)*, *Chroma*, dan *Mel Spectrogram* dapat dilakukan dengan lebih efisien. Hasil dari ekstraksi fitur tersebut kemudian digunakan sebagai data mentah untuk melatih model *Convolutional Neural network* (CNN) agar dapat mengklasifikasikan genre musik secara otomatis. Setiap prosedur dilakukan secara metodis dan terstruktur untuk memastikan keberhasilan pada tahap pembelajaran dan evaluasi model klasifikasi.

4.1.2 Augmentasi Data

Peneliti melakukan augmentasi data untuk menentukan jumlah dan variasi data yang tersedia sebelum melanjutkan ke tahap ekstraksi fitur dan pelatihan model. Augmentasi data adalah teknik yang biasa digunakan dalam Machine Learning, khususnya ketika *dataset* yang dimiliki relatif terbatas atau tidak seimbang. Augmentasi data bertujuan untuk memperluas variasi data latih yang tersedia, terutama mengingat keterbatasan jumlah data asli yang dimiliki.

Teknik augmentasi ini dilakukan dengan cara memodifikasi file audio yang sudah ada, seperti menambahkan noise, mengubah pitch, atau melakukan time-stretching. Setelah proses augmentasi selesai, data yang telah dimodifikasi disimpan sebagai file baru dan digabungkan dengan data asli. Dengan demikian, jumlah dataset bertambah secara signifikan dan variasinya juga lebih beragam.

Keberadaan data augmentasi ini memberikan kontribusi penting terhadap kualitas pelatihan model secara keseluruhan. Berdasarkan hasil pelatihan, model yang dilatih dengan data yang telah diaugmentasi mampu mencapai akurasi yang lebih stabil pada data uji, serta menunjukkan peningkatan performa dalam mengenali genre musik Pop Islami dan Pop secara lebih akurat dan konsisten.

```
# 2. Definisi semua augmentasi sebagai lambda yang pasti matching signature (y, sr)
AUGMENTATIONS = {
    'original': lambda y, sr: y,
    'noise' : lambda y, sr: y + 0.005 * np.random.randn(len(y)),
    'stretch': lambda y, sr: librosa.effects.time_stretch(y, rate=1.1),
    'pitch': lambda y, sr: librosa.effects.pitch_shift(y, sr=sr, n_steps=2),
    'shift': lambda y, sr: np.roll(y, int(np.random.uniform(-0.2, 0.2) * len(y)))
}
```

Gambar 4. 2 Augmentasi Data

Gambar 4.2 merupakan cuplikan kode program yang digunakan untuk mendefinisikan berbagai jenis teknik augmentasi data yang diterapkan terhadap sinyal audio. Augmentasi ini bertujuan untuk memperkaya variasi data latih agar model memiliki kemampuan generalisasi yang lebih baik dan mengurangi risiko *overfitting*. Seluruh metode augmentasi ditulis dalam bentuk fungsi lambda yang memiliki *signature* konsisten, yaitu *y* untuk data sinyal audio dan *sr* untuk *sampling rate*.

4.1.3 Ekstraksi Fitur Audio

Setelah proses augmentasi data selesai, langkah berikutnya adalah ekstraksi fitur audio. Ekstraksi fitur adalah tahap penting dalam pengolahan data audio karena fitur-fitur ini akan digunakan sebagai representasi numerik dari sinyal audio untuk dilatih ke dalam model *Machine Learning*.

Tiga jenis fitur suara yang diekstraksi dalam penelitian ini adalah *Mel Frequency Cepstral Coefficients* (MFCC), *Chroma Frequencies*, dan *Mel Spectrogram*. Fitur-fitur ini dipilih karena telah terbukti dapat menangkap karakteristik penting dari sinyal suara, terutama dalam hal pengenalan genre musik.

- a. MFCC digunakan untuk menangkap pola frekuensi berdasarkan persepsi pendengaran manusia, yang sangat relevan dalam membedakan jenis musik.
- b. *Chroma* merepresentasikan distribusi energi dalam 12 nada musik (kromatik), sehingga membantu mengenali struktur harmoni atau akord.
- c. *Mel Spectrogram* menampilkan representasi spektrum frekuensi yang telah diubah ke dalam skala mel, yang lebih sesuai dengan sensitivitas telinga manusia.

Proses ekstraksi dilakukan menggunakan pustaka **Librosa** pada Python. Setiap file audio hasil augmentasi diproses satu per satu untuk diubah menjadi gambar fitur (visualisasi), yang kemudian akan menjadi input bagi model CNN pada tahap pelatihan. Tahapan ini sangat penting karena kualitas fitur yang dihasilkan akan berpengaruh langsung terhadap kinerja model dalam melakukan klasifikasi genre. Dengan menggunakan kombinasi beberapa jenis fitur, diharapkan model dapat memahami beragam aspek dari sinyal audio secara lebih mendalam.

```
# Ekstraksi fitur
mfcc_feat = librosa.feature.mfcc(y=y_aug, sr=sr, n_mfcc=13)
chroma = librosa.feature.chroma_stft(y=y_aug, sr=sr)
mel = librosa.feature.melspectrogram(y=y_aug, sr=sr, n_mels=40)
mel_db = librosa.power_to_db(mel, ref=np.max)

mfcc_mean = mfcc_feat.mean(axis=1)
chroma_mean = chroma.mean(axis=1)
mel_mean = mel_db.mean(axis=1)

vector = np.hstack([mfcc_mean, chroma_mean, mel_mean])
```

Gambar 4. 3 ekstraksi fitur audio

Gambar 4.3 menampilkan potongan kode *Python* yang digunakan untuk mengekstraksi fitur dari sinyal audio menggunakan pustaka *Librosa*. Proses ekstraksi fitur ini merupakan tahap awal yang penting dalam sistem pengenalan suara atau klasifikasi audio. Dalam potongan kode tersebut, tiga jenis fitur utama diekstraksi, yaitu: MFCC diekstraksi menggunakan (`librosa.feature.mfcc`), *Chroma Feature* diekstraksi dengan (`librosa.feature.chroma_stft`), dan Mel-spectrogram dihasilkan melalui (`librosa.feature.melspectrogram`) dan dikonversi ke skala desibel (dB) menggunakan (`librosa.power_to_db`).

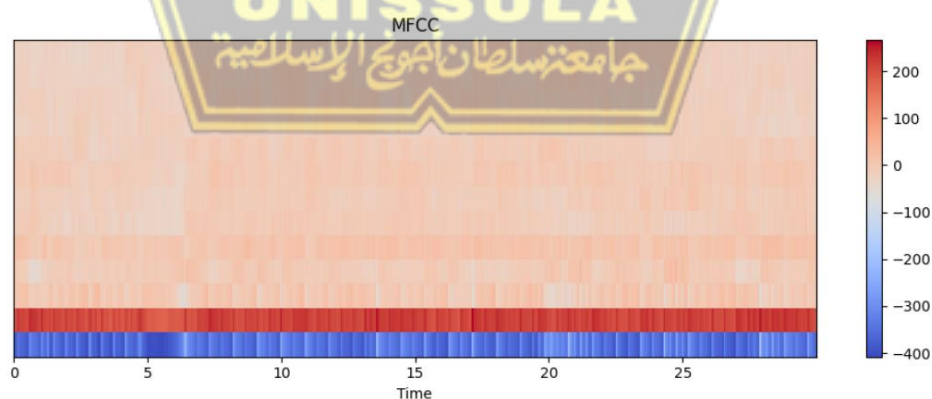
Setelah fitur-fitur tersebut dihitung, nilai mean dari masing-masing fitur dihitung sepanjang sumbu waktu (*axis=1*) untuk menghasilkan vektor fitur tetap. Ketiga mean tersebut kemudian digabungkan secara horizontal (*np.hstack*) ke dalam satu vektor fitur gabungan (*vector*), yang siap digunakan untuk proses selanjutnya, seperti pelatihan model klasifikasi.

4.1.4 Visualisasi Fitur Audio

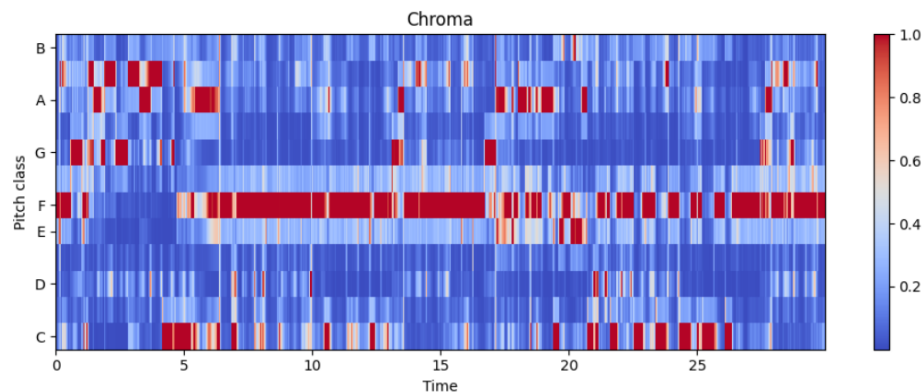
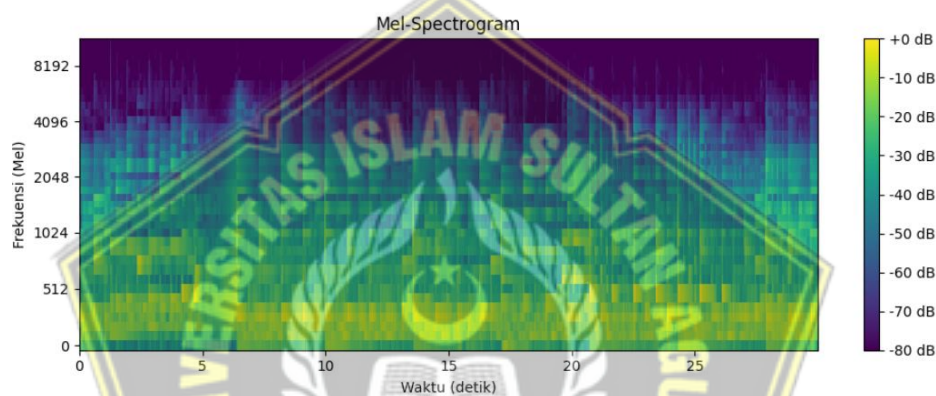
Setelah proses ekstraksi fitur dilakukan, langkah selanjutnya adalah memvisualisasikan fitur-fitur tersebut untuk memperoleh gambaran pola atau karakteristik dari sinyal audio. Visualisasi ini membantu dalam memahami bagaimana informasi frekuensi dan intensitas suara terdistribusi sepanjang waktu.

Tiga jenis fitur utama yang divisualisasikan meliputi:

1. *MFCC (Mel Frequency Cepstral Coefficients)* : Visualisasi MFCC menampilkan perubahan koefisien terhadap waktu dalam bentuk heatmap. Pola ini dapat menunjukkan struktur vokal atau perbedaan fonem dalam sinyal suara.
2. *Chroma Features* : representasi Chroma divisualisasikan sebagai peta spektral yang memperlihatkan intensitas nada-nada dasar (pitch class) dalam satu oktaf. Fitur ini berguna untuk mendeteksi harmoni dan progresi nada dalam audio musik.
3. *Mel-Spectrogram* : Mel-spectrogram menggambarkan distribusi energi dalam domain waktu-frekuensi berdasarkan skala Mel, yang lebih dekat dengan persepsi pendengaran manusia. Visualisasi ini dapat mengungkapkan informasi penting seperti ritme, intonasi, dan dinamika suara.



Gambar 4. 4 visualisasi MFCC

Gambar 4. 5 Visualisasi *Chroma*Gambar 4. 6 Visualisasi *Mel-Spectrogram*

Gambar 4.4 – gambar 4.6 merupakan hasil visualisasi fitur audio. Visualisasi MFCC menunjukkan bagaimana energi frekuensi suara terdistribusi dalam bentuk koefisien cepstral yang menyerupai persepsi pendengaran manusia terhadap suara. Warna pada grafik menunjukkan nilai intensitas dari masing-masing koefisien, yang menggambarkan bagaimana energi frekuensi tersusun dalam bentuk yang lebih kompak dan representative. Visualisasi *Chroma* menggambarkan intensitas 12 kelas *pitch* (nada) dari C hingga B, yang berguna untuk menangkap informasi harmonik dan tonal dari audio. Setiap baris menggambarkan kekuatan *pitch* tertentu, dan variasi warna menggambarkan intensitas relatif pada waktu tertentu. Sementara itu, *Mel-Spectrogram* memberikan representasi energi frekuensi dalam skala Mel terhadap waktu, yang memperlihatkan bagaimana energi suara tersebar di berbagai frekuensi selama durasi audio. Warna dalam grafik menunjukkan intensitas energi suara pada setiap kombinasi waktu dan frekuensi. Representasi ini

sangat berguna dalam identifikasi pola frekuensi yang berulang atau signifikan dalam sinyal suara.

4.1.5 Persiapan Data dan Pembagian Dataset

Pada tahap persiapan data ini, langkah pertama yang dilakukan adalah mengubah label dari format kategorikal (biasanya berupa teks) menjadi format numerik menggunakan metode *label encoding*. Hal ini sangat penting karena sebagian besar algoritma machine learning hanya dapat memproses data dalam bentuk angka. Label encoding memberikan representasi numerik unik untuk setiap kelas, sehingga model dapat memahami dan memproses target secara efektif.

Setelah label berhasil diubah, dataset kemudian dibagi menjadi tiga bagian utama: *training set*, *validation set*, dan *test set*. Pembagian ini bertujuan untuk memastikan bahwa proses pelatihan dan evaluasi model berlangsung secara objektif dan optimal.

1. *Training set* (70%) digunakan untuk melatih model agar dapat mengenali pola dan hubungan dalam data. Model menggunakan data ini untuk belajar dan menyesuaikan parameternya.
2. *Validation set* (15%) berfungsi sebagai data pengujian selama proses pelatihan, yang membantu dalam melakukan tuning hyperparameter dan memilih model terbaik tanpa mengganggu data test. Data ini memungkinkan penilaian performa model secara real-time untuk mencegah overfitting.
3. *Test set* (15%) adalah data yang digunakan hanya setelah proses pelatihan dan validasi selesai, untuk mengukur performa akhir model dengan data yang benar-benar baru dan belum pernah dilihat sebelumnya. Ini memberikan gambaran objektif tentang kemampuan model saat diterapkan pada data nyata.

Pembagian dataset dilakukan secara *stratified*, yang berarti proporsi masing-masing kelas dalam label tetap terjaga pada setiap subset (*training*, *validation*, dan *test*). Dengan demikian, distribusi kelas yang seimbang dapat dipertahankan, menghindari bias dalam proses pelatihan dan evaluasi.

```

# Encode label ke angka
le = LabelEncoder()
y_encoded = le.fit_transform(labels)
X = np.array(features)
np.save('label_encoder.npy', le.classes_)

# Split: train 70%, val 15%, test 15%
from sklearn.model_selection import train_test_split
X_train, X_temp, y_train, y_temp = train_test_split(X, y_encoded, test_size=0.3, stratify=y_encoded, random_state=42)
X_val, X_test, y_val, y_test = train_test_split(X_temp, y_temp, test_size=0.5, stratify=y_temp, random_state=42)

```

Gambar 4. 7 Potongan kode untuk konversi label dan *split dataset*

Gambar 4.7 merupakan potongan kode yang menunjukkan proses encoding label dan pembagian dataset. Label yang semula berupa format string dikonversi ke format numerik menggunakan *LabelEncoder* agar dapat diproses oleh model. Selanjutnya, dataset dibagi menjadi tiga bagian utama, yaitu data pelatihan (70%), data validasi (15%), dan data pengujian (15%). Pembagian dilakukan secara stratified untuk menjaga distribusi kelas yang seimbang di setiap subset data, sehingga model dapat belajar secara adil dan terhindar dari bias terhadap kelas tertentu.

Setelah proses pembagian dataset dilakukan, dilakukan pencetakan jumlah data pada masing-masing subset untuk memastikan bahwa pembagian telah sesuai dengan rasio yang diharapkan.

```

print("Total sampel :", X.shape[0])
print("    • Train    :", X_train.shape[0])
print("    • Val      :", X_val.shape[0])
print("    • Test     :", X_test.shape[0])

```

Total sampel : 2000
 • Train : 1400
 • Val : 300
 • Test : 300

Gambar 4. 8 Distribusi jumlah data pelatihan, validasi, dan pengujian

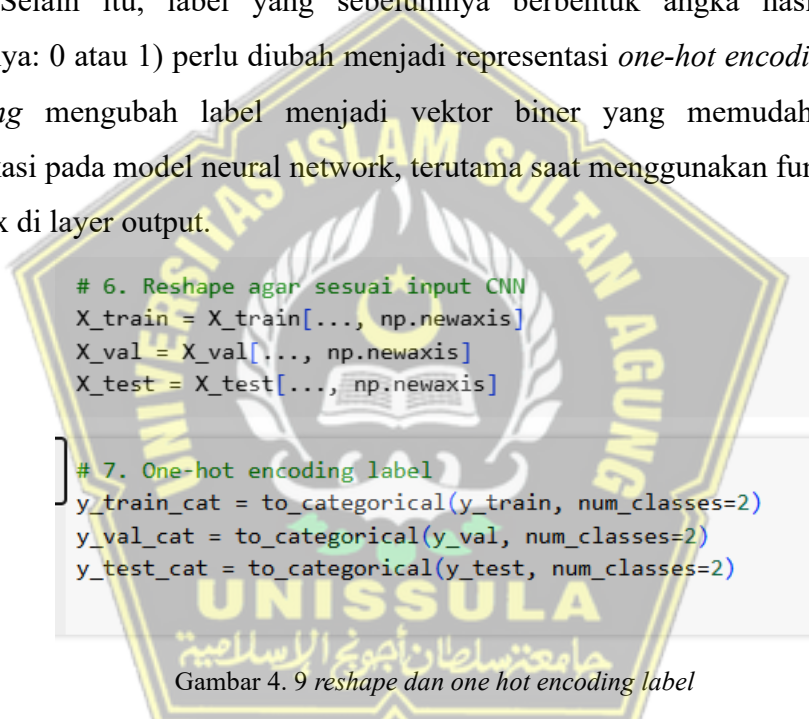
Gambar 4.8 merupakan *Output* jumlah data setelah pembagian dataset. Total terdapat 1.000 sampel, yang merupakan hasil dari proses augmentasi data sebelumnya. Dataset dibagi secara stratified dengan rasio 70% data pelatihan (1400 sampel), 15% data validasi (300 sampel), dan 15% data pengujian (300 sampel). Pembagian *stratified* menjaga proporsi kelas agar tetap seimbang pada masing-masing subset, sehingga mengurangi risiko bias selama pelatihan dan evaluasi model.

4.1.6 Persiapan Data untuk Model CNN

Setelah dataset dibagi menjadi data pelatihan, validasi, dan pengujian, langkah berikutnya adalah melakukan penyesuaian bentuk (*reshape*) dan representasi label agar sesuai dengan kebutuhan arsitektur model CNN.

CNN (*Convolutional Neural Network*) biasanya memerlukan input dengan format tiga dimensi: tinggi, lebar, dan jumlah kanal (*channels*). Karena data input kita awalnya hanya berdimensi dua, perlu ditambahkan satu dimensi tambahan agar sesuai format tersebut.

Selain itu, label yang sebelumnya berbentuk angka hasil encoding (misalnya: 0 atau 1) perlu diubah menjadi representasi *one-hot encoding*. *One-hot encoding* mengubah label menjadi vektor biner yang memudahkan proses klasifikasi pada model neural network, terutama saat menggunakan fungsi aktivasi softmax di layer output.



```
# 6. Reshape agar sesuai input CNN
X_train = X_train[..., np.newaxis]
X_val = X_val[..., np.newaxis]
X_test = X_test[..., np.newaxis]

# 7. One-hot encoding label
y_train_cat = to_categorical(y_train, num_classes=2)
y_val_cat = to_categorical(y_val, num_classes=2)
y_test_cat = to_categorical(y_test, num_classes=2)
```

Gambar 4. 9 *reshape* dan *one hot encoding label*

Gambar 4.9 menunjukkan potongan kode proses penyesuaian bentuk data input (*reshape*) dan konversi label ke bentuk one-hot encoding. Penyesuaian ini dilakukan agar data sesuai dengan format input CNN dan label dapat diproses oleh fungsi aktivasi softmax.

4.1.7 Pemilihan Model

Pada tahap ini, peneliti memilih arsitektur model yang digunakan untuk melakukan klasifikasi genre musik Pop Islami dan Pop. Berdasarkan karakteristik data berupa fitur audio dua dimensi (seperti MFCC, *Chroma*, dan *Mel Spectrogram*), model yang dipilih adalah ***Convolutional Neural network (CNN)***

karena kemampuannya dalam mengenali pola spasial dan temporal dari data audio yang telah direpresentasikan sebagai citra atau urutan angka.

Model CNN dinilai sesuai karena telah banyak digunakan dalam tugas-tugas klasifikasi audio dan citra, serta memiliki struktur yang mendukung pengenalan pola fitur kompleks dari data. CNN dapat mengurangi kompleksitas data input melalui proses konvolusi dan pooling, yang membuat model lebih efisien dan cepat dalam belajar fitur penting.

Model CNN dibangun dengan menggunakan pustaka Keras. Lapisan pertama adalah *Conv1D*, yang berfungsi untuk mengekstrak fitur lokal dari data input. Lapisan pooling (*MaxPooling1D*) digunakan untuk mengurangi dimensi data dan mempercepat proses pelatihan. Beberapa lapisan *Dropout* ditambahkan untuk mencegah overfitting. Setelah itu, data diratakan (*Flatten*) dan diproses oleh lapisan *Dense* sebagai fully connected layer. Output akhir terdiri dari dua neuron dengan fungsi aktivasi *softmax*, yang mewakili dua kelas genre: Islami dan Pop.

Pemilihan arsitektur dan parameter CNN ini didasarkan pada pertimbangan efisiensi, kesesuaian dengan jenis data, serta referensi dari penelitian sebelumnya yang berhasil menggunakan CNN dalam klasifikasi audio.

Model: "sequential"

Layer (type)	Output Shape	Param #
conv1d (Conv1D)	(None, 63, 32)	128
max_pooling1d (MaxPooling1D)	(None, 31, 32)	0
dropout (Dropout)	(None, 31, 32)	0
conv1d_1 (Conv1D)	(None, 29, 64)	6,208
max_pooling1d_1 (MaxPooling1D)	(None, 14, 64)	0
dropout_1 (Dropout)	(None, 14, 64)	0
flatten (Flatten)	(None, 896)	0
dense (Dense)	(None, 64)	57,408
dropout_2 (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 2)	130

Total params: 63,874 (249.51 KB)

Trainable params: 63,874 (249.51 KB)

Non-trainable params: 0 (0.00 B)

Gambar 4. 10 output model cnn

Gambar 4.7 merupakan hasil dari perintah `model.summary()` yang menampilkan arsitektur jaringan *Convolutional Neural network* (CNN) yang digunakan dalam penelitian ini. Model ini disusun secara sekuensial, terdiri dari beberapa lapisan, antara lain:

- a. Dua buah lapisan Conv1D yang bertugas mengekstraksi pola lokal dari data input, masing-masing menghasilkan 32 dan 64 filter.
- b. Lapisan MaxPooling1D yang digunakan setelah setiap Conv1D untuk mereduksi dimensi data dan mempercepat proses pelatihan.
- c. Lapisan Dropout yang ditambahkan setelah pooling layer dan dense layer untuk mengurangi risiko overfitting dengan cara menonaktifkan sebagian neuron selama pelatihan.
- d. Lapisan Flatten yang mengubah data multidimensi menjadi vektor satu dimensi.
- e. Lapisan Dense (*fully connected*) sebanyak dua kali; yang pertama berisi 64 neuron dan yang terakhir berisi 2 neuron sebagai output layer (karena klasifikasi dilakukan untuk dua kelas: Islami dan Pop).
- f. Aktivasi akhir menggunakan fungsi *softmax*, karena klasifikasi bersifat multi-class (dua kelas).

Model ini memiliki total parameter sebanyak 63.874, yang semuanya bersifat trainable, artinya dapat dioptimalkan selama proses pelatihan.

4.1.8 Pelatihan Model

Setelah model CNN selesai dibangun, langkah selanjutnya adalah melakukan proses pelatihan (*training*) terhadap model menggunakan dataset yang telah melalui proses ekstraksi fitur. Pelatihan ini bertujuan untuk mengoptimalkan bobot-bobot pada jaringan agar model dapat mengenali pola dari data dan melakukan klasifikasi genre musik Islami dan Pop secara akurat.

```
# 9. Latih model
history = model.fit(X_train, y_train_cat, validation_data=(X_val, y_val_cat), epochs=30, batch_size=16)
```

Gambar 4. 11 latih model

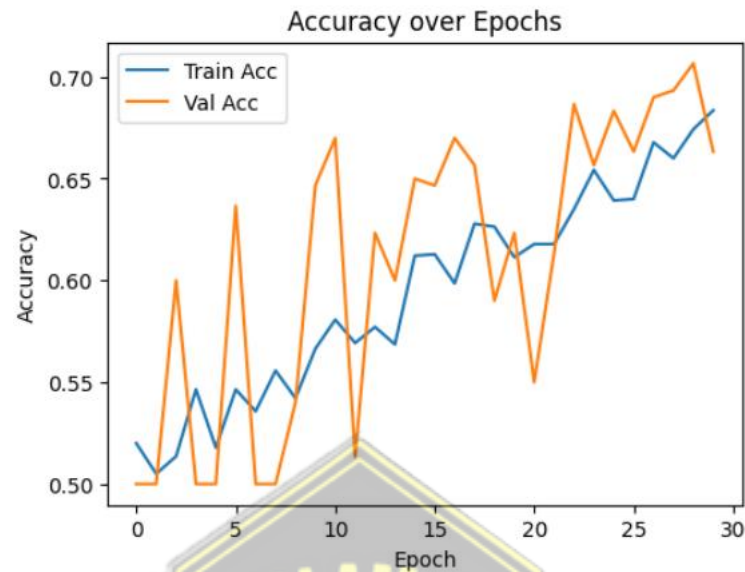
Gambar 4.11 merupakan tampilan kode program yang digunakan untuk melatih model *Convolutional Neural network* (CNN). Pada gambar ini ditunjukkan proses pelatihan model dengan menggunakan data latih dan data validasi yang telah melalui proses ekstraksi fitur. Parameter seperti jumlah epoch dan ukuran batch ditentukan untuk mengatur lamanya proses pelatihan serta jumlah data yang diproses dalam setiap iterasi.

Kode pada gambar ini berfungsi untuk menjalankan proses pembelajaran model terhadap pola-pola yang terdapat pada data, serta mengevaluasi performanya pada data validasi di setiap *epoch*. Hasil dari pelatihan ini akan digunakan sebagai dasar dalam tahap evaluasi dan pengujian model klasifikasi genre musik Islami dan Pop.

4.1.9 Evaluasi Model

Setelah proses pelatihan selesai, tahap selanjutnya adalah mengevaluasi performa model *Convolutional Neural network* (CNN) dalam melakukan klasifikasi genre musik Islami dan Pop. Evaluasi ini bertujuan untuk mengukur seberapa baik model dalam mengenali pola dari data uji dan sejauh mana generalisasi model terhadap data baru.

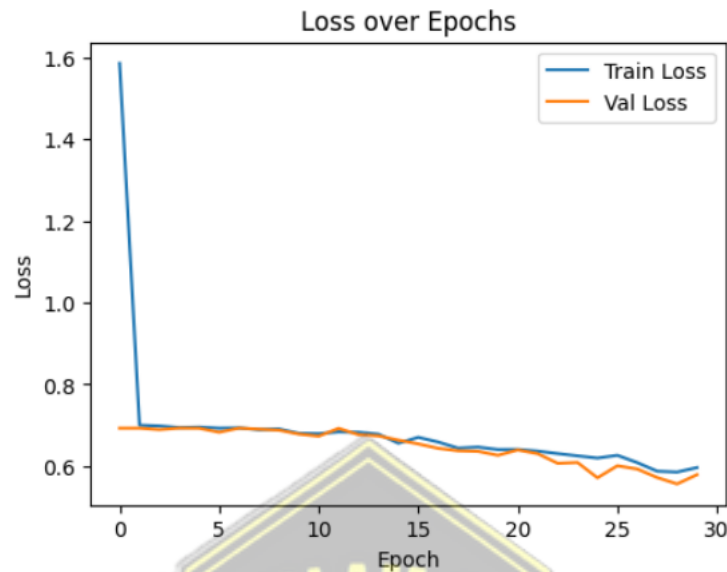
Evaluasi dilakukan dengan menggunakan data validasi yang sebelumnya telah dipisahkan dari data pelatihan. Beberapa metrik yang digunakan antara lain adalah akurasi, nilai *loss*, *recall*, *f1-score*, dan *confusion matrix*. Akurasi menunjukkan persentase prediksi yang benar terhadap total data uji, sedangkan *loss* menggambarkan seberapa jauh prediksi model dari label yang sebenarnya. *Confusion matrix* memberikan informasi lebih detail mengenai jumlah prediksi benar dan salah pada masing-masing kelas.



Gambar 4. 12 grafik nilai akurasi

Gambar 4.12 Gambar di atas menunjukkan grafik Accuracy over Epochs yang menggambarkan akurasi model pada data pelatihan (Train Acc) dan data validasi (Val Acc) selama 30 epoch pelatihan. Grafik ini memberikan gambaran mengenai bagaimana kinerja model meningkat seiring berjalannya proses training.

Terlihat bahwa akurasi model pada data pelatihan meningkat secara bertahap dari sekitar 0.51 pada epoch ke-0 hingga mencapai sekitar 0.69 pada epoch ke-30. Sementara itu, akurasi pada data validasi terlihat berfluktuasi lebih tajam, dimulai dari 0.50 dan sempat mengalami beberapa penurunan drastis, bahkan ke titik 0.0 pada beberapa epoch awal. Namun, seiring bertambahnya epoch, akurasi validasi menunjukkan tren kenaikan dan stabil di kisaran 0.68–0.72 pada epoch akhir. Perbedaan fluktuasi antara akurasi pelatihan dan validasi ini menandakan bahwa model mungkin mengalami tantangan dalam generalisasi pada awal pelatihan, tetapi secara keseluruhan, grafik ini menunjukkan tren positif yang mengindikasikan bahwa model belajar dengan baik dan menunjukkan kemampuan generalisasi yang cukup pada data yang tidak terlihat sebelumnya. Meskipun terdapat lonjakan dan penurunan tajam di awal, pada akhirnya model mampu mencapai akurasi validasi maksimum sekitar 0.72, yang lebih tinggi dibanding akurasi pelatihan akhir.



Gambar 4. 13 grafik nilai loss

Gambar 4.13 memperlihatkan grafik Loss over Epochs, yaitu visualisasi perubahan nilai loss selama proses pelatihan model dalam 30 epoch. Grafik ini membandingkan nilai Train Loss (garis biru) dengan Validation Loss (garis oranye), yang menunjukkan sejauh mana kesalahan prediksi model terhadap data pelatihan dan validasi.

Pada awal pelatihan (epoch ke-0), train loss sangat tinggi, yaitu sekitar 1.6, menandakan bahwa model awalnya memiliki kesalahan prediksi yang besar terhadap data pelatihan. Namun, pada epoch berikutnya, train loss langsung menurun drastis dan mendekati nilai validation loss, yaitu sekitar 0.7. Setelah itu, baik train loss maupun validation loss cenderung menurun perlahan secara stabil hingga akhir pelatihan, masing-masing mencapai sekitar 0.59 dan 0.57 pada epoch ke-30. Polanya menunjukkan bahwa model berhasil belajar dari data tanpa mengalami overfitting yang signifikan. Hal ini terlihat dari jarak antara train loss dan validation loss yang relatif kecil sepanjang epoch, serta kecenderungan keduanya untuk menurun bersamaan. Dengan demikian, model menunjukkan performa yang baik dalam hal generalisasi dan kestabilan selama pelatihan.

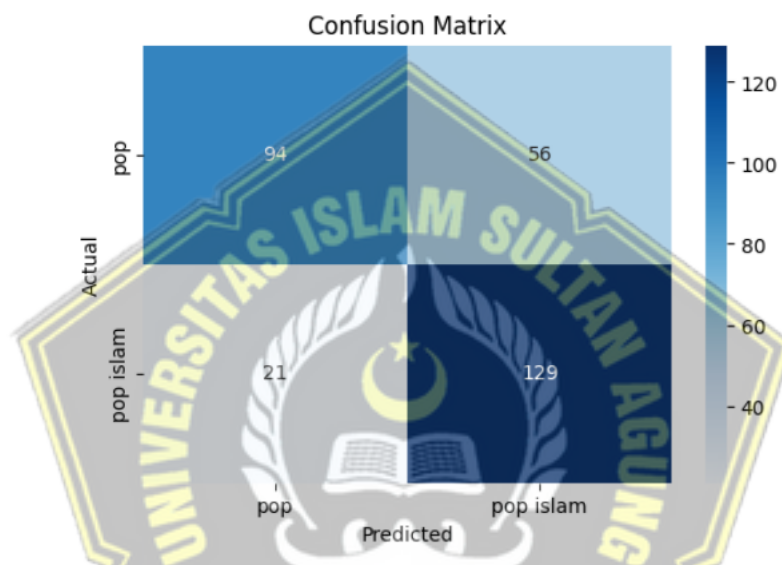
Tabel 4. 1 Tabel *Classification report*

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	<i>Support</i>
Pop	0.82	0.63	0.71	150
Pop Islami	0.70	0.86	0.77	150
<i>Accuracy</i>			0.74	300
<i>Macro Avg</i>	0.76	0.74	0.74	300
<i>Weighted Avg</i>	0.76	0.74	0.74	300

Berdasarkan laporan klasifikasi pada gambar di atas, model menunjukkan performa yang cukup seimbang namun masih memiliki ruang untuk peningkatan. Untuk kelas Pop, nilai precision yang diperoleh adalah 0.82, menunjukkan bahwa sebagian besar prediksi Pop memang benar. Namun, nilai recall-nya hanya 0.63, yang berarti masih banyak lagu Pop yang tidak berhasil dikenali sebagai Pop. Sebaliknya, pada kelas Pop Islam, model berhasil mencapai recall yang cukup tinggi yaitu 0.86, menandakan bahwa model cukup andal dalam mendeteksi lagu-lagu Islami. Namun, precision-nya lebih rendah, yaitu 0.70, yang mengindikasikan bahwa masih terdapat kesalahan model dalam memprediksi lagu-lagu non-Islami sebagai Islami.

Secara keseluruhan, model memperoleh nilai accuracy sebesar 0.74, dengan nilai rata-rata f1-score (baik macro maupun weighted average) yang sama yaitu 0.74. Ini menunjukkan bahwa performa model relatif stabil antar kelas, namun ketidakseimbangan antara precision dan recall pada masing-masing kelas tetap menjadi perhatian.

Setelah menganalisis metrik evaluasi pada classification report, diperlukan visualisasi yang dapat memperjelas performa model dalam mengenali masing-masing kelas secara lebih konkret. Untuk itu, digunakan *Confusion Matrix* sebagai alat bantu visual yang menunjukkan jumlah prediksi benar dan salah yang dilakukan oleh model terhadap data uji pada masing-masing kelas. *Confusion Matrix* ini akan membantu mengetahui kecenderungan model dalam salah mengenali suatu kelas dan melihat distribusi kesalahan prediksi secara lebih jelas.



Gambar 4. 14 *confusion matrix* hasil prediksi model

Gambar 4.14 menunjukkan hasil dari confusion matrix hasil prediksi model terhadap data uji pada dua kelas, yaitu Pop Islami dan Pop. Confusion matrix ini digunakan untuk mengevaluasi performa klasifikasi model secara lebih detail, tidak hanya berdasarkan akurasi, tetapi juga melihat jumlah kesalahan dalam masing-masing kelas.

Berdasarkan gambar tersebut, diperoleh bahwa:

- Sebanyak 129 data dengan label asli Pop Islami berhasil diklasifikasikan dengan benar sebagai Islami (*True Positive*)
- Sebanyak 21 data Pop Islami diklasifikasikan keliru sebagai Pop (*False Negative*)
- Sebanyak 94 data Pop berhasil diklasifikasikan dengan benar sebagai Pop (*True Negative*)

- d. Namun terdapat 56 data Pop yang salah diklasifikasikan sebagai Islami (*False Positive*).

Dari confusion matrix ini dapat dilihat bahwa model memiliki performa yang lebih baik dalam mengklasifikasikan kelas *Pop Islami* dibandingkan kelas *Pop*. Hal ini terlihat dari nilai True Positive yang tinggi untuk kelas Islami (129), namun model masih mengalami kesulitan dalam mengenali data kelas Pop dengan benar, yang tercermin dari jumlah *False Positive* yang cukup besar (56) yang artinya banyak lagu *Pop* yang justru diklasifikasikan sebagai *Pop Islam*.

Kondisi ini dapat disebabkan oleh beberapa faktor seperti distribusi data yang tidak seimbang antara kedua kelas, atau fitur-fitur yang digunakan belum cukup membedakan karakteristik khas dari lagu *Pop*.

4.1.10 Implementasi Model

Setelah model CNN selesai dilatih, tahap selanjutnya adalah mengimplementasikan model tersebut untuk melakukan klasifikasi genre musik terhadap data uji yang baru. Implementasi dilakukan dengan memanfaatkan model yang telah disimpan dalam format *.h5*, kemudian dimuat kembali ke dalam sistem untuk digunakan dalam proses prediksi.

Pengujian dilakukan dengan mengambil file audio baru dari luar data latih, kemudian melalui proses yang sama seperti sebelumnya, yaitu ekstraksi fitur audio menggunakan MFCC, Chroma, dan Mel Spectrogram. Hasil ekstraksi fitur kemudian dikonversi ke dalam bentuk array dan dijadikan input untuk model.

Pada tahap implementasi, sistem klasifikasi genre musik yang telah dikembangkan diintegrasikan ke dalam aplikasi antarmuka berbasis web menggunakan *Streamlit*. Streamlit dipilih karena mampu membangun aplikasi web secara interaktif dan ringan tanpa memerlukan pengaturan backend yang kompleks. Aplikasi ini bertugas menerima masukan berupa file audio dari pengguna, kemudian melakukan prediksi genre menggunakan model CNN yang telah dilatih.

Namun, karena Streamlit tidak dapat dijalankan langsung pada lingkungan Google Colab, maka diperlukan solusi tambahan berupa *ngrok*, yaitu layanan tunneling yang memungkinkan akses publik terhadap aplikasi Streamlit yang dijalankan secara lokal.

```
!pip install -q streamlit pyngrok
!ngrok authtoken 2xDLCyYJq78WuzeIKCctU830W0Y_5kJeQgoz4QnFAskjr4ven
```

44.3/44.3 kB	2.8 MB/s	eta 0:00:00
9.9/9.9 MB	39.7 MB/s	eta 0:00:00
6.9/6.9 MB	49.8 MB/s	eta 0:00:00
79.1/79.1 kB	6.5 MB/s	eta 0:00:00

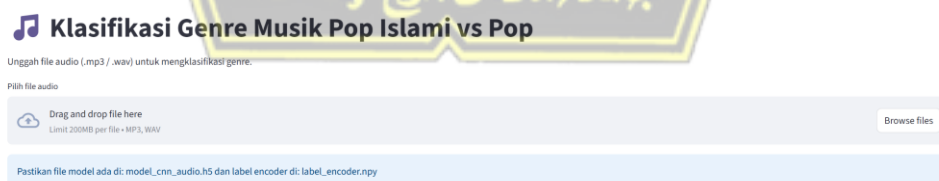
Authtoken saved to configuration file: /root/.config/ngrok/ngrok.yml

Gambar 4. 15 penginstalasian ngrok

Gambar 4.15 menunjukkan proses instalasi library *streamlit* dan *pyngrok*, serta konfigurasi authtoken dari Ngrok pada lingkungan kerja berbasis Python. Ngrok digunakan sebagai solusi tunneling yang memungkinkan aplikasi *Streamlit* yang berjalan secara lokal (*localhost*) dapat diakses secara publik melalui URL yang disediakan oleh Ngrok. Hal ini sangat bermanfaat dalam proses deployment aplikasi berbasis web yang dibangun menggunakan Streamlit di lingkungan cloud seperti Google Colab.

Pada gambar tersebut, perintah `!pip install -q streamlit pyngrok` digunakan untuk menginstal kedua library yang dibutuhkan. Setelah proses instalasi berhasil, perintah `!ngrok authtoken <token>` digunakan untuk mengatur authtoken agar Ngrok dapat diakses secara sah melalui akun pengguna yang terdaftar.

Sistem ini diberi nama "**Klasifikasi Genre Musik Pop Islami dan Pop**", sesuai dengan fungsi utamanya dalam membedakan dua kategori musik berdasarkan fitur audio.



Gambar 4. 16 tampilan awal sistem klasifikasi genre musik Islami dan Pop

Gambar 4.16 menunjukkan antarmuka aplikasi web yang dibangun menggunakan Streamlit untuk melakukan klasifikasi genre musik antara genre Islami dan Pop. Aplikasi ini memungkinkan pengguna untuk mengunggah file audio dengan ekstensi .mp3, atau .wav guna diprediksi genrenya oleh model klasifikasi yang telah dilatih sebelumnya. Pada bagian atas antarmuka, terdapat judul “Klasifikasi Genre Musik Pop Islami vs Pop” yang menggambarkan fungsi utama dari aplikasi. Di bawahnya terdapat komponen **file uploader** yang memungkinkan pengguna untuk menyeret dan meletakkan file atau memilih file audio dari perangkat mereka.

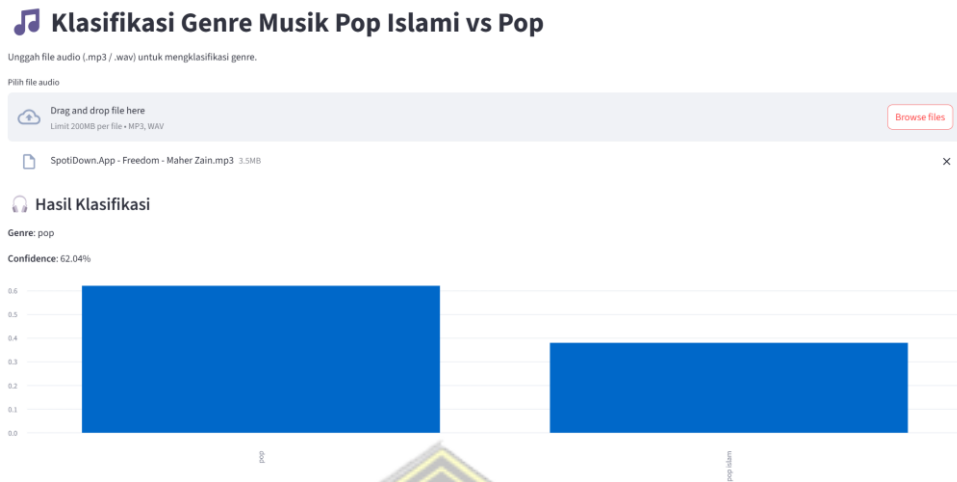
Selain itu, aplikasi menampilkan pesan peringatan berwarna biru yang mengingatkan pengguna untuk memastikan bahwa file model (model_cnn_audio.h5) dan file label encoder (label_encoder.npy) telah tersedia di direktori yang sesuai agar proses prediksi dapat berjalan dengan lancar.



Gambar 4. 17 tampilan file audio di proses

Gambar 4.17 menunjukkan tampilan antarmuka aplikasi klasifikasi genre musik setelah pengguna mengunggah file audio dan sistem sedang memproses file tersebut. Aplikasi ini dibangun menggunakan Streamlit dan digunakan untuk mengklasifikasikan genre lagu ke dalam dua kategori, yaitu genre Pop Islami dan Pop.

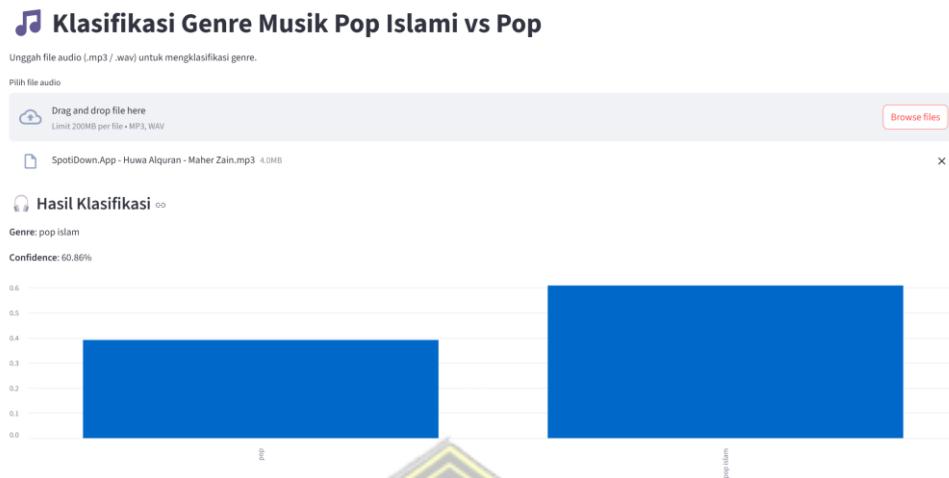
Pada bagian tengah antarmuka terlihat file audio dengan nama “Freedom - Maher ...” sedang diproses, ditandai dengan adanya progress bar berwarna biru. Hal ini menandakan bahwa model klasifikasi tengah menganalisis fitur dari file audio tersebut untuk menentukan hasil klasifikasi genre.



Gambar 4. 18 tampilan hasil klasifikasi genre pop

Gambar 4.18 memperlihatkan tampilan antarmuka aplikasi setelah file audio berhasil diproses dan sistem menyajikan hasil klasifikasi genre lagu. Proses ini merupakan lanjutan dari tahapan pada Gambar 4.17, di mana file audio sedang dianalisis oleh model klasifikasi.

Pada bagian “Hasil Klasifikasi”, ditampilkan bahwa lagu yang diunggah dalam hal ini *Freedom* oleh Maher Zain — di klasifikasi masuk dalam kategori genre Pop dengan tingkat keyakinan (confidence) sebesar 62.04%. Angka ini menunjukkan seberapa yakin model terhadap prediksinya terhadap genre yang dipilih. Di bawahnya, ditampilkan grafik batang (*bar chart*) yang memperlihatkan distribusi probabilitas terhadap masing-masing label genre. Terdapat dua batang yang mewakili dua kelas, yaitu Islami dan Pop. Pada grafik batang Islami nilai distribusi sebesar 37.96% dan pada grafik batang Pop nilai distribusi sebesar 62.04%.



Gambar 4. 19 Tampilan hasil klasifikasi genre Islami

Gambar 4.18 menunjukkan antarmuka aplikasi klasifikasi genre musik setelah pengguna mengunggah file audio berjudul “Huwa Alquran – Maher Zain.mp3”. Berdasarkan hasil prediksi yang ditampilkan, model mengklasifikasikan lagu tersebut sebagai genre Islami dengan tingkat kepercayaan (confidence) sebesar 60.86%. Grafik batang di bawahnya menggambarkan distribusi probabilitas dari dua kelas yang tersedia yaitu Islami sekitar 60.86% dan Pop sekitar 39.14%

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

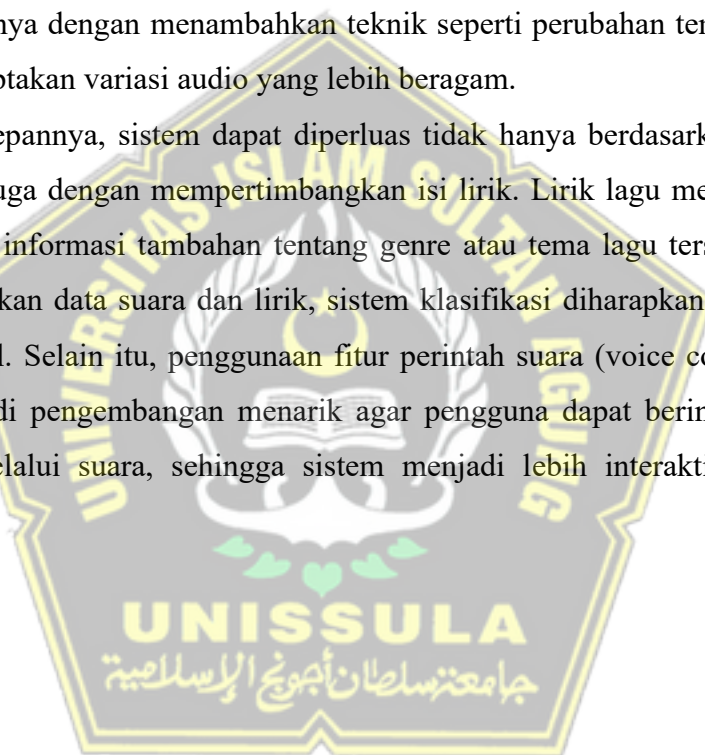
Berdasarkan hasil yang telah dicapai dalam penelitian ini, dapat disimpulkan bahwa teknik *Deep Learning*, khususnya metode *Convolutional Neural Network* (CNN), dapat digunakan secara efektif untuk membangun model klasifikasi otomatis yang mampu membedakan genre musik Islami dan Pop. Proses pengembangan dimulai dengan pengumpulan data audio dari *Spotify*, dilanjutkan dengan augmentasi untuk meningkatkan variasi data, serta ekstraksi fitur audio menggunakan MFCC, *Chroma*, dan *Mel Spectrogram* yang dikonversi ke dalam bentuk citra untuk keperluan pelatihan model.

Selanjutnya, evaluasi model menunjukkan bahwa model CNN yang telah dilatih mampu melakukan klasifikasi dengan tingkat akurasi sebesar 0.74 (74%), dengan nilai *precision* 0.70 (Pop Islami) dan 0.82 (Pop), *recall* 0.86 (Pop Islami) dan 0.63 (Pop), serta *f1-score* sebesar 0.77 (Pop Islami) dan 0.71 (Pop). Hasil ini menunjukkan bahwa model cukup baik dalam mengenali genre Islami, namun masih memiliki keterbatasan dalam mengenali genre Pop secara konsisten. Hal ini kemungkinan besar dipengaruhi oleh keterbatasan jumlah data, ketidakseimbangan distribusi genre, serta kompleksitas karakteristik audio Pop yang lebih bervariasi. Hasil ini menunjukkan bahwa pendekatan CNN memiliki kemampuan dalam mempelajari pola-pola mendasar dari fitur audio, dan dapat menjadi pijakan awal yang solid dalam pengembangan sistem klasifikasi genre musik Islami secara otomatis di masa mendatang. Dengan peningkatan pada jumlah data, strategi preprocessing, dan arsitektur model, performa sistem diharapkan dapat terus ditingkatkan hingga mencapai tingkat akurasi yang lebih tinggi dan stabil.

5.2 Saran

Berdasarkan hasil yang diperoleh dari penelitian ini, terdapat beberapa hal yang dapat dijadikan saran untuk pengembangan lebih lanjut. Pertama, jumlah data yang digunakan masih tergolong terbatas. Maka dari itu, disarankan untuk menambah jumlah dan variasi lagu dari masing-masing genre, serta mempertimbangkan penambahan genre lain. Hal ini bertujuan agar model memiliki cakupan klasifikasi yang lebih luas dan mampu mengenali lebih banyak jenis musik secara akurat. Selain itu, proses augmentasi data juga dapat dikembangkan lebih lanjut, misalnya dengan menambahkan teknik seperti perubahan tempo atau nada untuk menciptakan variasi audio yang lebih beragam.

Ke depannya, sistem dapat diperluas tidak hanya berdasarkan suara lagu saja, tetapi juga dengan mempertimbangkan isi lirik. Lirik lagu memiliki potensi memberikan informasi tambahan tentang genre atau tema lagu tersebut. Dengan menggabungkan data suara dan lirik, sistem klasifikasi diharapkan dapat bekerja lebih optimal. Selain itu, penggunaan fitur perintah suara (voice command) juga dapat menjadi pengembangan menarik agar pengguna dapat berinteraksi secara langsung melalui suara, sehingga sistem menjadi lebih interaktif dan mudah digunakan.



DAFTAR PUSTAKA

- Ayu, G., Giri, V. M., & Radhitya, L. (2023). KLASIFIKASI GENRE MUSIK MENGGUNAKAN TEKNIK PEMBELAJARAN MESIN. *Jurnal Teknologi Informasi dan Komputer*, 9(1).
- Brawijaya, U., Hani Haqyah, S., & Henryranu Prasetio, B. (2017). *Fakultas Ilmu Komputer SISTEM KLASIFIKASI GENRE MUSIK BERDASARKAN RITME DAN FREKUENSI MENGGUNAKAN EKSTRAKSI FITUR MFCC DAN F0 DENGAN ALGORITMA DECISION TREE* (Vol. 1, Nomor 1). <http://j-ptiik.ub.ac.id>
- Fardhani, S. M., Wihardi, Y., & Piantari, E. (2021). *Klasifikasi Genre Musik Dengan Mel Frequency Cepstral Coefficient Dan Spektogram Menggunakan Convolutional Neural Network* (Vol. 4, Nomor 1). <https://ejournal.upi.edu/index.php/JATIKOM>
- Guo, Y. (2024). Research on music genre recognition method based on deep learning. *Molecular & Cellular Biomechanics*, 21(1), 373. <https://doi.org/10.62617/mcb.v21i1.373>
- Khamitkar, S. D., & Lokhande, S. N. (2021). ADVANCEMENTS IN AUDIO SIGNAL PROCESSING USING PYTHON: A COMPREHENSIVE REVIEW OF LIBRARIES AND APPLICATIONS. *International Research Journal of Modernization in Engineering Technology and Science* www.irjmets.com @International Research Journal of Modernization in Engineering, 2582–5208. <https://doi.org/10.56726/IRJMETS18040>
- Kumar, M. K., Sujanasri, K., Neha, B., Akshara, G., Chugh, P., & Haindavi, P. (2023). Automated Music Genre Classification through Deep Learning Techniques. *E3S Web of Conferences*, 430. <https://doi.org/10.1051/e3sconf/202343001033>
- Kurniawan Soeganda, R., & Budiyantri, N. (2024). PRO DAN KONTRA MUSIK DALAM PERSPEKTIF ISLAM. Dalam *Jurnal Ilmiah Multidisiplin Terpadu* (Vol. 8, Nomor 6). <https://www.detaktangsel.com/opini/pro-dan-kontra-musik-dalam-kajian-fikih-dan-bagaimana->
- Ma, X. (2023). Music Genre Classification Based on Machine Learning Methods. Dalam *Highlights in Science, Engineering and Technology CSIC* (Vol. 2022).

- McFee, B., Raffel, C., Liang, D., Ellis, D., McVicar, M., Battenberg, E., & Nieto, O. (2015). librosa: Audio and Music Signal Analysis in Python. *Proceedings of the 14th Python in Science Conference*, 18–24. <https://doi.org/10.25080/majora-7b98e3ed-003>
- Pratama, A. (2021). *KLASIFIKASI GENRE MUSIK POPULER MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK DENGAN DATA AUGMENTATION*.
- Singh, A., Chodankar, S., & Suvarna, A. (2021). AUDIO FEATURE EXTRACTION TOOLS. Dalam *International Research Journal of Modernization in Engineering Technology and Science* www.irjmets.com @International Research Journal of Modernization in Engineering. www.irjmets.com
- Suhartini, Wijaya, L. K., & Pratiwi, N. A. (2020). Penerapan Algoritma K-Means Untuk Pendataan Obat Berdasarkan Laporan Bulanan Pada Dinas Kesehatan Kabupaten Lombok Timur. *Jurnal Informatika dan Teknologi*, 3(2), 147–156.
- Wairata, C. R., Swedia, E. R., & Cahyanti, M. (2021). PENGKLASIFIKASIAN GENRE MUSIK INDONESIA MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK. *Sebatik*, 25(1), 255–261. <https://doi.org/10.46984/sebatik.v25i1.1286>
- Yahya, & Mahpuz. (2019). Penggunaan Algoritma K-Means Untuk Menganalisis Pelanggan Potensial Pada Dealer SPS. *Jurnal Informatika dan Teknologi*, 2(2), 109–118.