

**KLASIFIKASI USIA DAN GENDER BERBASIS SUARA
MENGUNAKAN ALGORITMA HYBRID CNN-BILSTM-
TRANSFORMER**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

M.NASYID YUNITIAN RIZAL

22.11.5073

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**KLASIFIKASI USIA DAN GENDER BERBASIS SUARA
MENGUNAKAN ALGORITMA HYBRID CNN-BILSTM-
TRANSFORMER**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

M.NASYID YUNITIAN RIZAL

22.11.5073

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**KLASIFIKASI USIA DAN GENDER BERBASIS SUARA
MENGUNAKAN ALGORITMA HYBRID CNN-BILSTM-
TRANSFORMER**

yang disusun dan diajukan oleh

M.Nasyid Yunitian Rizal

22.11.5073

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 2 Maret 2026

Dosen Pembimbing,



Theopilus Bayu Sasongko, S.Kom., M.Eng.

NIK. 190302375

HALAMAN PENGESAHAN
SKRIPSI
KLASIFIKASI USIA DAN GENDER BERBASIS SUARA
MENGGUNAKAN ALGORITMA HYBRID CNN-BILSTM-
TRANSFORMER

yang disusun dan diajukan oleh

M.Nasyid Yunitian Rizal

22.11.5073

Telah dipertahankan di depan Dewan Penguji
pada tanggal 2 Maret 2026

Susunan Dewan Penguji

Nama Penguji

Nuri Cahyono, M.Kom.
NIK. 190302278

Mujivanto, M.Kom
NIK. 190302492

Theopilus Bayu Sasongko, S.Kom., M.Eng.
NIK. 190302375

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 2 Maret 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusrini, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : M.Nasyid Yunitian Rizal
NIM : 22.11.5073

Menyatakan bahwa Skripsi dengan judul berikut:

Klasifikasi Usia dan Gender Berbasis Suara Menggunakan Algoritma Hybrid CNN-BiLSTM-Transformer

Dosen Pembimbing : Theopilus Bayu Sasongko, S.Kom., M.Eng.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 2 Maret 2026

Yang Menyatakan,



M.Nasyid Yunitian Rizal

HALAMAN PERSEMBAHAN

Puji syukur ke hadirat Allah SWT atas segala rahmat dan karunia-Nya sehingga penyusunan skripsi ini dapat diselesaikan dengan baik. Dengan segala kerendahan hati dan rasa syukur, karya sederhana ini penulis persembahkan kepada:

1. Kedua orang tua tercinta, khususnya Bapak Yunianto dan Ibu Siti Baroroh, atas segala doa yang tidak pernah putus, pengorbanan, serta dukungan moril maupun materiel yang telah mengiringi setiap langkah penulis selama masa studi.
2. Kakak tercinta, Farda Yunitia Dewi, atas segala dukungan, motivasi, serta kepedulian yang senantiasa diberikan kepada penulis hingga skripsi ini selesai.
3. Dosen Pembimbing Skripsi, Bapak Theopilus Bayu Sasongko, S.Kom., M.Eng., atas segala bimbingan, kesabaran, waktu, dan arahan berharga yang telah menuntun penulis dari awal perancangan hingga selesainya penelitian ini.
4. Jajaran Dosen Penguji, Bapak Nuri Cahyono, M.Kom. dan Bapak Mujiyanto, M.Kom., atas segala masukan, kritik kritis, dan evaluasi yang sangat membangun demi kesempurnaan metodologi penelitian ini.
5. Rekan-rekan seperjuangan kelas Informatika 9 (IF 9) beserta sahabat dan teman-teman terdekat, atas solidaritas, semangat, dukungan diskusi, dan bantuan yang telah banyak mewarnai dinamika masa perkuliahan penulis.
6. Almamater tercinta, Program Studi Informatika, Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta, sebagai wadah penulis untuk menimba ilmu dan mengembangkan diri.

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT atas segala limpahan rahmat, karunia, dan hidayah-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul "Klasifikasi Usia dan Gender Berbasis Suara Menggunakan Algoritma Hybrid CNN-BiLSTM-Transformer" dengan lancar. Skripsi ini disusun untuk memenuhi salah satu syarat dalam mencapai derajat Sarjana Komputer pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta.

Penyelesaian skripsi ini tentu tidak lepas dari dukungan, doa, serta bantuan dari berbagai pihak. Oleh karena itu, melalui kesempatan ini penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada Bapak Theopilus Bayu Sasongko, S.Kom., M.Eng., selaku dosen pembimbing skripsi yang senantiasa meluangkan waktu, tenaga, dan pikirannya untuk memberikan bimbingan, arahan, maupun masukan yang sangat berarti dari awal hingga selesainya penelitian ini. Ucapan terima kasih juga penulis tujukan kepada jajaran dosen penguji atas segala evaluasi dan kritik konstruktif yang diberikan demi kesempurnaan metodologi dalam karya ini, serta kepada seluruh tenaga pendidik di lingkungan Program Studi Informatika Universitas AMIKOM Yogyakarta yang telah membekali penulis dengan berbagai ilmu pengetahuan yang bermanfaat selama masa perkuliahan.

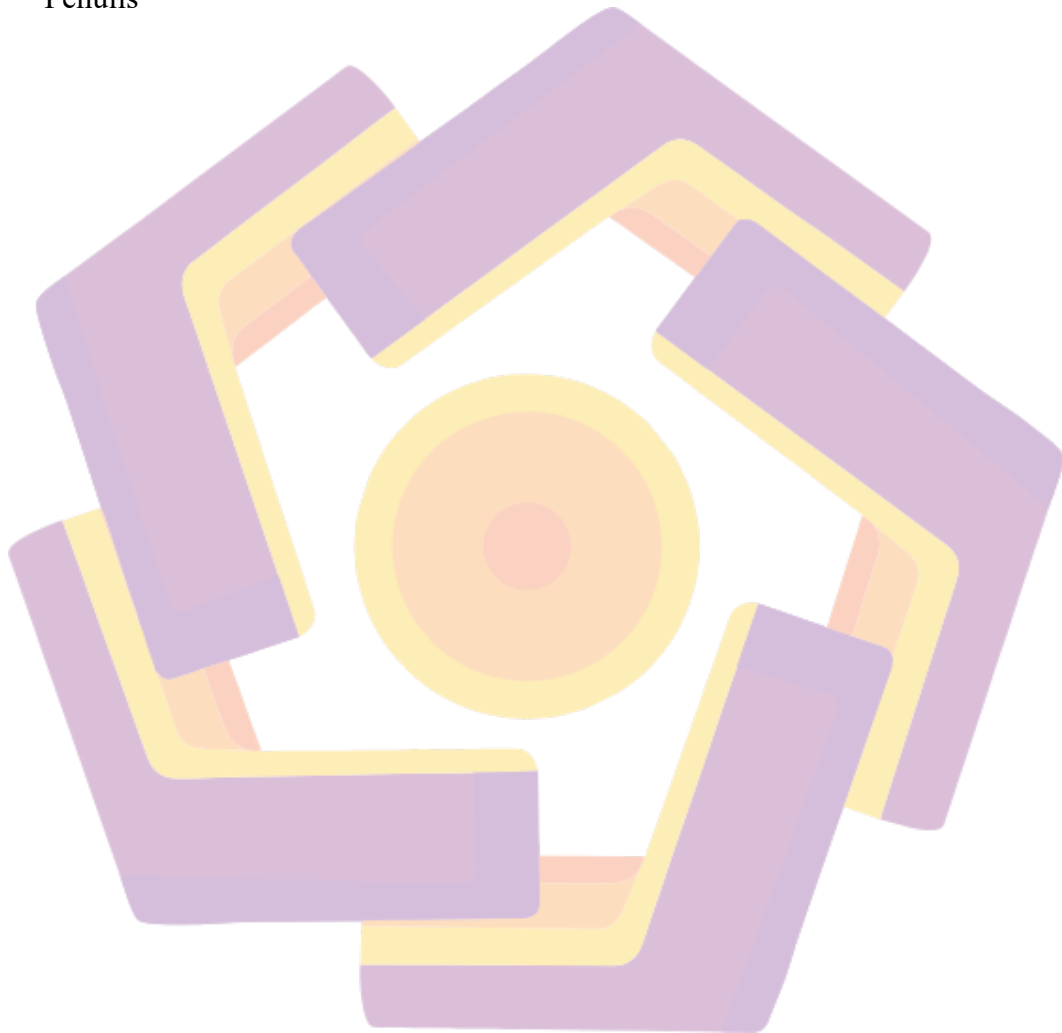
Penghargaan dan rasa terima kasih yang tak terhingga penulis haturkan kepada kedua orang tua beserta keluarga besar atas segala curahan kasih sayang, doa yang tidak pernah putus, serta dukungan penuh baik secara moril maupun materiel yang senantiasa menjadi sumber kekuatan utama bagi penulis. Tidak lupa, penulis juga mengapresiasi seluruh rekan seperjuangan serta berbagai pihak yang tidak dapat disebutkan satu per satu, namun telah banyak memberikan bantuan dan kemudahan dalam penyusunan tugas akhir ini.

Penulis menyadari sepenuhnya bahwa penyusunan laporan skripsi ini masih memiliki berbagai kekurangan karena keterbatasan ilmu dan pengalaman yang dimiliki. Oleh sebab itu, segala bentuk kritik dan saran yang membangun

sangat penulis harapkan. Akhir kata, semoga karya tulis ini dapat memberikan manfaat yang nyata dan menjadi referensi yang berguna bagi pengembangan ilmu pengetahuan, khususnya di bidang kecerdasan buatan.

Yogyakarta, 2 Maret 2026

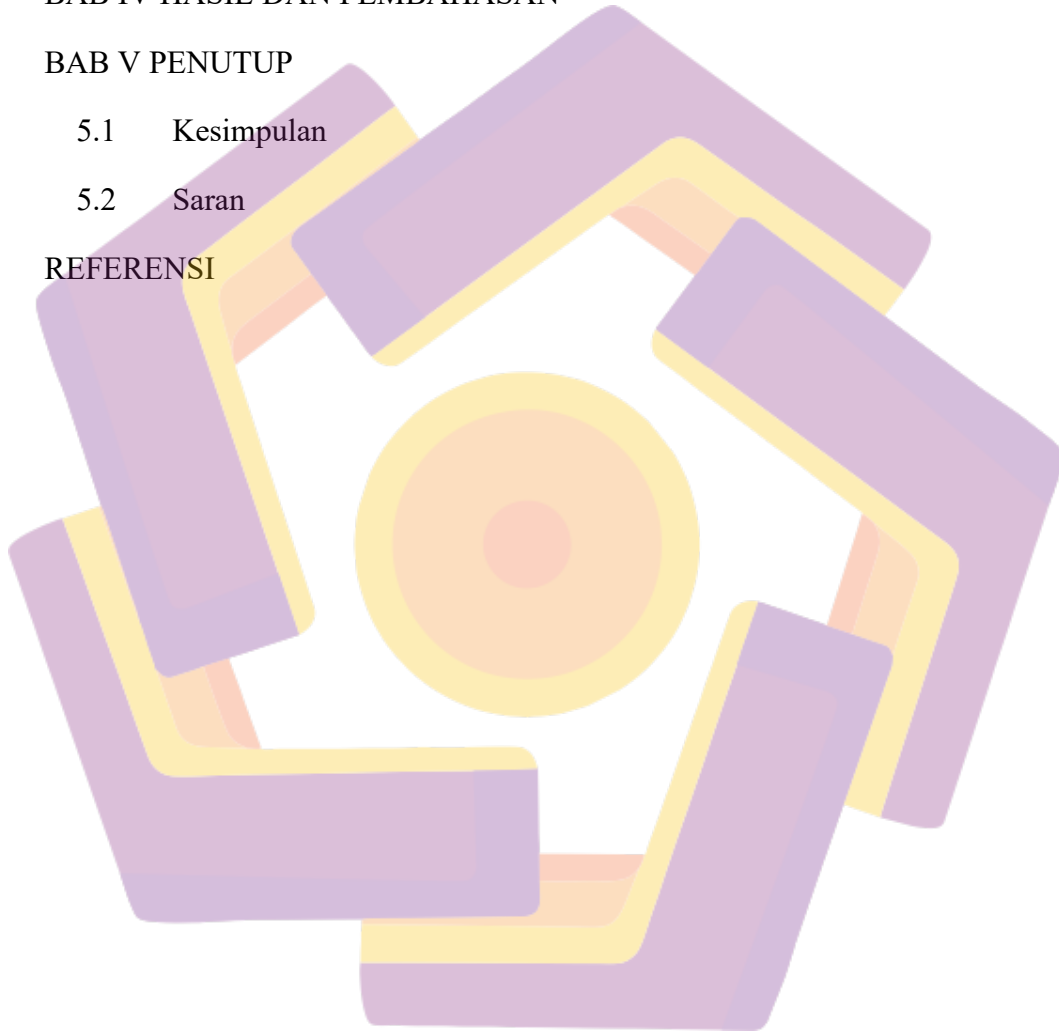
Penulis



DAFTAR ISI

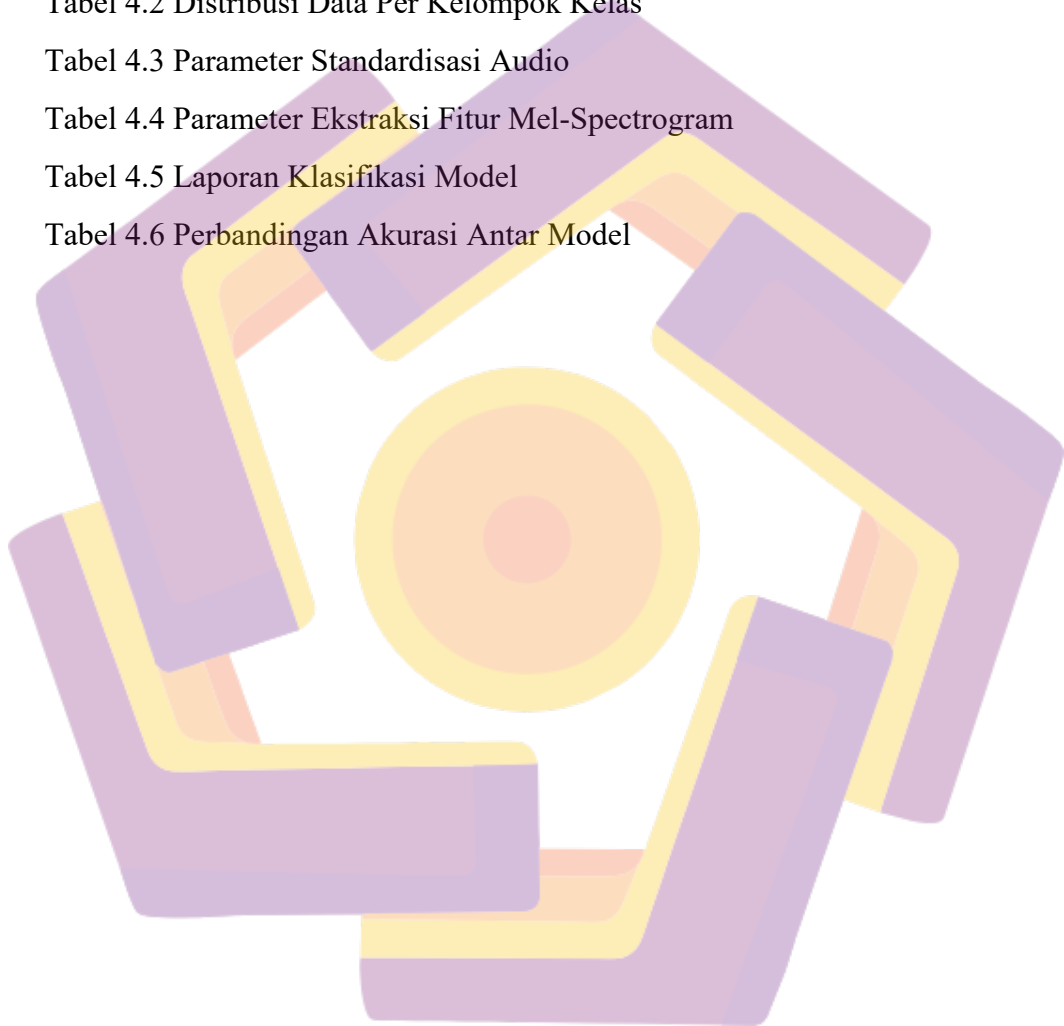
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH	xiv
INTISARI	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7
2.1 Studi Literatur	7
2.2 Dasar Teori	15

BAB III METODE PENELITIAN	32
3.1 Objek Penelitian	32
3.2 Alur Penelitian	32
3.3 Alat dan Bahan	37
BAB IV HASIL DAN PEMBAHASAN	40
BAB V PENUTUP	57
5.1 Kesimpulan	57
5.2 Saran	57
REFERENSI	59



DAFTAR TABEL

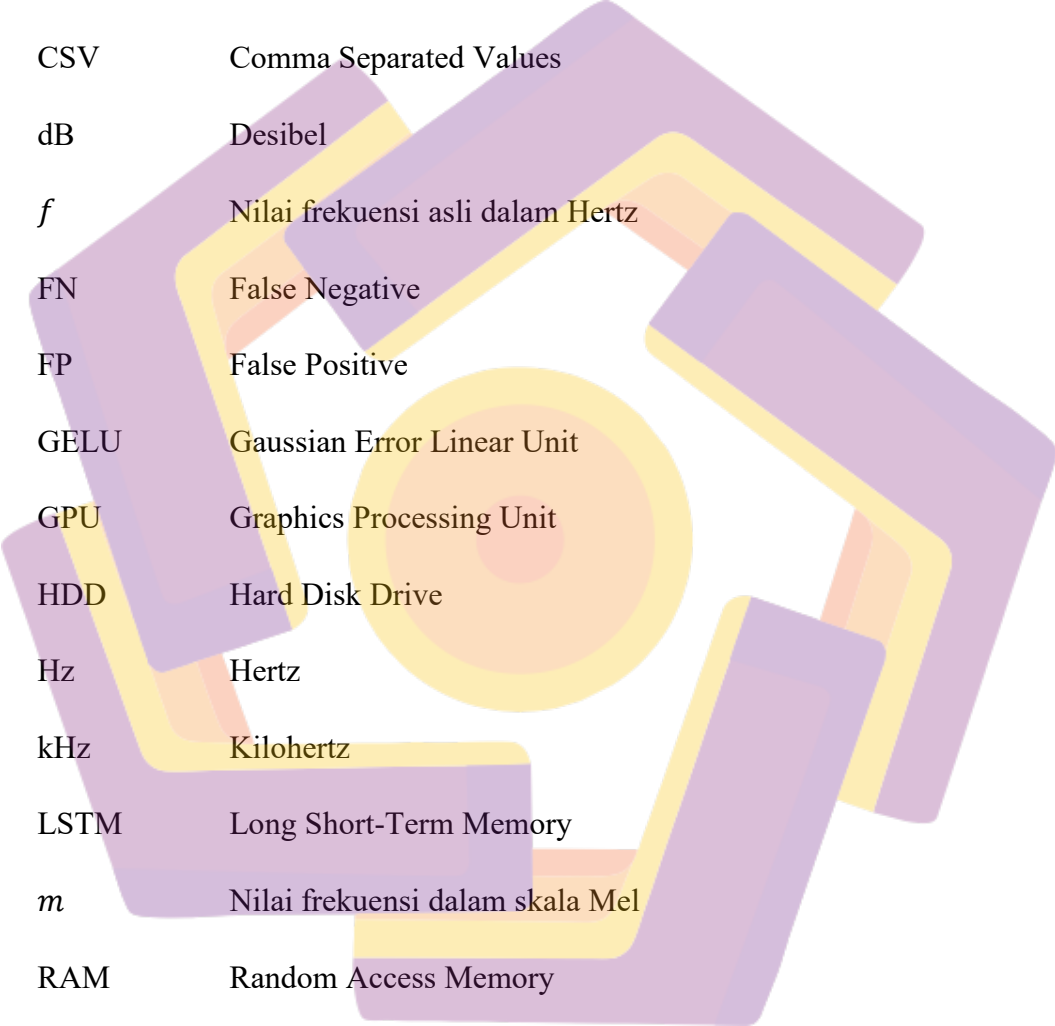
Tabel 2.1 Keaslian Penelitian	11
Tabel 3.1 Deskripsi Atribut Dataset Common Voice	38
Tabel 4.1 Sampel Dataset Common Voice	41
Tabel 4.2 Distribusi Data Per Kelompok Kelas	43
Tabel 4.3 Parameter Standardisasi Audio	44
Tabel 4.4 Parameter Ekstraksi Fitur Mel-Spectrogram	46
Tabel 4.5 Laporan Klasifikasi Model	53
Tabel 4.6 Perbandingan Akurasi Antar Model	54



DAFTAR GAMBAR

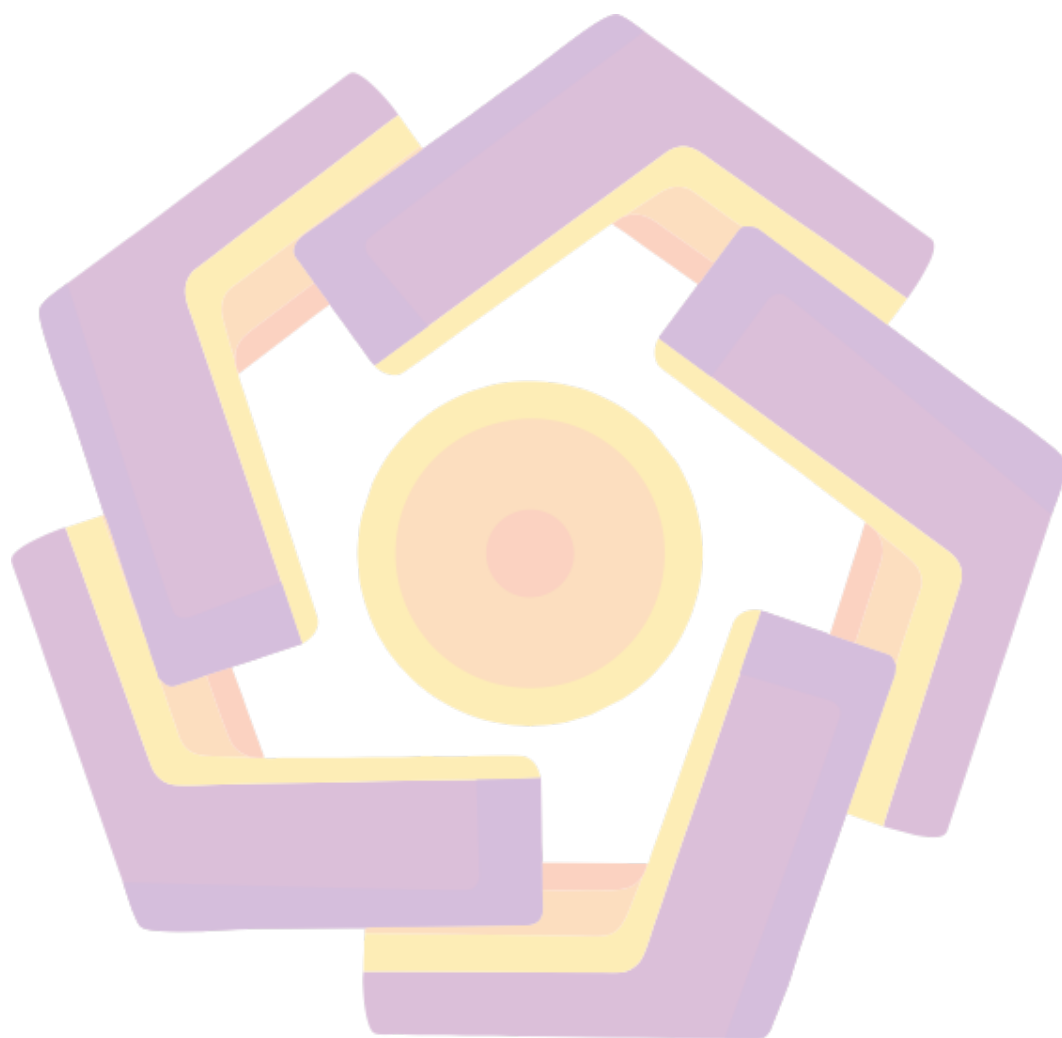
Gambar 2.1 Visualisasi Mel-Spectrogram	17
Gambar 2.2 Arsitektur Umum CNN	19
Gambar 2.3 Arsitektur RNN	20
Gambar 2.4 Arsitektur LSTM	21
Gambar 2.5 Arsitektur Bi-LSTM	23
Gambar 2.6 Arsitektur Transformer	24
Gambar 2.7 Interpretasi Confusion Matrix	28
Gambar 3.1 Alur Penelitian	33
Gambar 4.1 Hasil Penyimpanan Fitur Ekstraksi dalam Format NumPy	47
Gambar 4.2 Diagram Alur Data dari Input hingga Output	48
Gambar 4.3 Grafik Akurasi Pelatihan dan Validasi Model Hibrida	51
Gambar 4.4 Confusion Matrix Hasil Prediksi Model Hibrida	52
Gambar 4.5 Tampilan Antarmuka Aplikasi Streamlit	55

DAFTAR LAMBANG DAN SINGKATAN



Adam	Adaptive Moment Estimation
Bi-LSTM	Bi-Directional Long Short-Term Memory
CNN	Convolutional Neural Network
CSV	Comma Separated Values
dB	Desibel
f	Nilai frekuensi asli dalam Hertz
FN	False Negative
FP	False Positive
GELU	Gaussian Error Linear Unit
GPU	Graphics Processing Unit
HDD	Hard Disk Drive
Hz	Hertz
kHz	Kilohertz
LSTM	Long Short-Term Memory
m	Nilai frekuensi dalam skala Mel
RAM	Random Access Memory
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
SE-Block	Squeeze-and-Excitation Block
SSD	Solid State Drive
STFT	Short-Time Fourier Transform

TN	True Negative
TP	True Positive
VRAM	Video Random Access Memory



DAFTAR ISTILAH



Accuracy	Rasio total prediksi benar terhadap keseluruhan data.
Bi-LSTM	LSTM yang memproses urutan data dua arah maju-mundur.
CNN	Jaringan saraf untuk mengekstraksi fitur visual atau spasial.
Confusion Matrix	Tabel perbandingan antara prediksi model dan label aktual.
Deep Learning	Pembelajaran mesin menggunakan arsitektur berlapis.
Epoch	Satu siklus pelatihan penuh pada seluruh dataset.
F1-Score	Rata-rata harmonik penyeimbang presisi dan recall.
GELU	Fungsi aktivasi untuk meningkatkan stabilitas model.
Loss Function	Fungsi pengukur kesalahan antara prediksi dan target.
Mel-Spectrogram	Representasi visual suara dalam skala frekuensi Mel.
Multi-Head Attention	Mekanisme fokus pada berbagai bagian data secara paralel.
Overfitting	Kondisi model hafal data latih tapi gagal pada data baru.
Precision	Tingkat ketepatan prediksi positif model.
Recall	Kemampuan model menemukan seluruh data positif.
Softmax	Fungsi pengubah output numerik menjadi probabilitas.
Transformer	Arsitektur berbasis atensi untuk menangkap konteks global.
Undersampling	Pengurangan data mayoritas agar seimbang.

INTISARI

Pengenalan karakteristik demografis pengguna, seperti usia dan gender, melalui suara merupakan tantangan krusial dalam pengembangan sistem interaksi manusia dan mesin yang lebih personal dan aman. Namun, penggunaan model *Deep Learning* arsitektur tunggal seringkali memiliki keterbatasan, di mana model CNN kurang optimal dalam menangkap urutan waktu, sedangkan model sekuensial seperti RNN sering kehilangan detail spasial pada data suara yang kompleks. Penelitian ini bertujuan untuk mengatasi masalah tersebut dengan mengembangkan model *Hybrid* yang mengintegrasikan *Convolutional Neural Network* (CNN), *Bidirectional Long Short-Term Memory* (Bi-LSTM), dan *Transformer*. Metode penelitian diawali dengan pengolahan dataset *Common Voice* menjadi representasi *Mel-Spectrogram*, penerapan teknik standarisasi audio, serta perancangan arsitektur gabungan di mana CNN mengekstraksi fitur visual spektral, Bi-LSTM menangkap konteks temporal dua arah, dan *Transformer* memfokuskan atensi pada fitur global yang dominan. Hasil pengujian membuktikan bahwa model usulan mampu mencapai akurasi akhir sebesar 91%, unggul dibandingkan model tunggal Bi-LSTM (74%) dan CNN standar (84%). Keberhasilan ini menegaskan bahwa sinergi ketiga algoritma tersebut merupakan pendekatan yang efektif untuk mengatasi variabilitas karakteristik suara manusia yang beragam. Implementasi sistem ini diharapkan dapat dimanfaatkan secara luas untuk meningkatkan keandalan sistem keamanan biometrik, personalisasi layanan *e-commerce*, serta alat bantu investigasi forensik digital.

Kata kunci: Klasifikasi Suara, *Deep Learning*, Arsitektur *Hybrid*, Usia dan Gender, *Mel-Spectrogram*.

ABSTRACT

Recognizing user demographic characteristics, such as age and gender, through voice is a crucial challenge in developing more personal and secure human-machine interaction systems. However, the use of single-architecture Deep Learning models often has limitations, where CNN models are less optimal in capturing temporal sequences, while sequential models like RNN often lose spatial details in complex voice data. This study aims to address this issue by developing a Hybrid model that integrates Convolutional Neural Network (CNN), Bidirectional Long Short-Term Memory (Bi-LSTM), and Transformer. The research method begins with processing the Common Voice dataset into Mel-Spectrogram representations, applying audio standardization techniques, and designing a combined architecture where CNN extracts spectral visual features, Bi-LSTM captures bidirectional temporal context, and Transformer focuses attention on dominant global features. The test results prove that the proposed model achieved a final accuracy of 91%, outperforming the single Bi-LSTM model (74%) and standard CNN (84%). This success confirms that the synergy of these three algorithms is an effective approach to overcome the variability of diverse human voice characteristics. The implementation of this system is expected to be widely utilized to improve the reliability of biometric security systems, e-commerce service personalization, and digital forensic investigation tools.

Keywords: Voice Classification, Deep Learning, Hybrid Architecture, Age and Gender, Mel-Spectrogram.