

**Optimasi Deteksi Aksen Daerah pada Ucapan Bahasa
Indonesia Menggunakan CNN dan Augmentasi Audio**

**SKRIPSI SARJANA REKAYASA TEKNOLOGI
INFORMATIKA**

Oleh

Dwi Tunjung Nawang Wulan

217064516084



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA**

UNIVERSITAS NASIONAL

2026

**Optimasi Deteksi Aksen Daerah pada Ucapan Bahasa
Indonesia Menggunakan CNN dan Augmentasi Audio**

SKRIPSI SARJANA

Karya ilmiah sebagai salah satu syarat untuk memperoleh gelar
Sarjana Teknologi Informatika dari Fakultas Teknologi Komunikasi
dan Informatika

Oleh

Dwi Tunjung Nawang Wulan

217064516084



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA
UNIVERSITAS NASIONAL**

2026

HALAMAN PENGESAHAN

TUGAS AKHIR

**Optimasi Deteksi Aksen Daerah pada Ucapan Bahasa Indonesia
Menggunakan CNN dan Augmentasi Audio**



Dwi Tunjung Nawang Wulan

217064516084

Dosen Pembimbing

A handwritten signature in black ink, consisting of stylized, overlapping letters and lines.

Ratih Titi Komala Sari, ST., MM., MMSI.

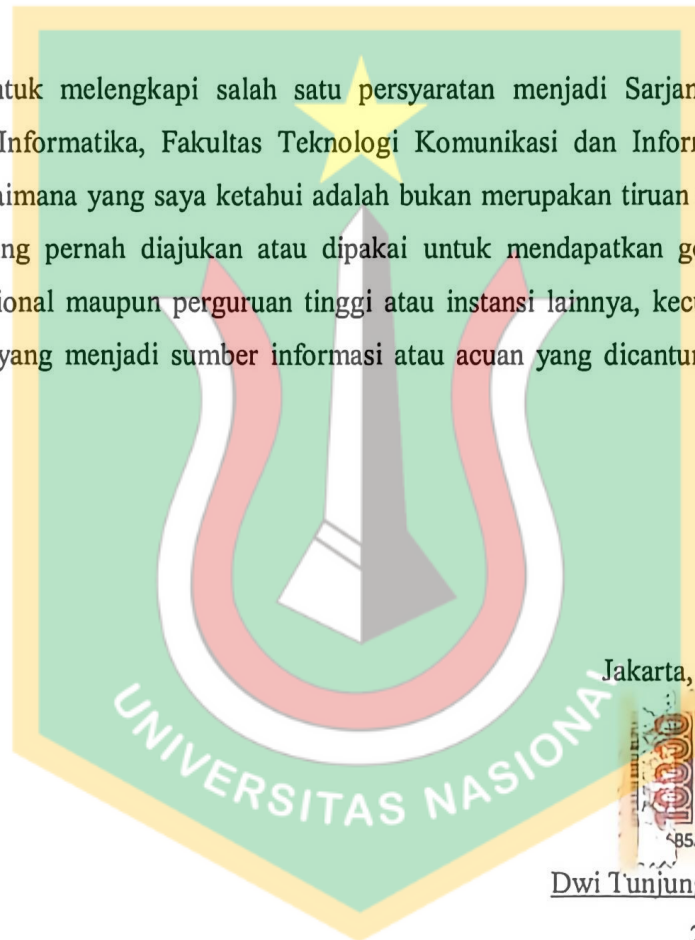
NIDN: 0301038302

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

Optimasi Deteksi Aksen Daerah pada Ucapan Bahasa Indonesia Menggunakan CNN dan Augmentasi Audio

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, sebagaimana yang saya ketahui adalah bukan merupakan tiruan atau publikasi dari Tugas Akhir yang pernah diajukan atau dipakai untuk mendapatkan gelar di lingkungan Universitas Nasional maupun perguruan tinggi atau instansi lainnya, kecuali pada bagian – bagian tertentu yang menjadi sumber informasi atau acuan yang dicantumkan sebagaimana mestinya.



Jakarta, 2 Maret 2026



Dwi Tunjung Nawang Wulan

217064516084

LEMBAR PERSETUJUAN REVIEW AKHIR

Tugas Akhir dengan judul :

Optimasi Deteksi Aksen Daerah pada Ucapan Bahasa Indonesia Menggunakan CNN dan Augmentasi Audio

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Review Akhir Semester Ganjil 2021-2022 pada tanggal (isi tanggal) Tahun 2026



LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Dwi Tunjung Nawang Wulan
NPM : 217064516084
Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
Program Studi : Informatika
Tanggal Sidang : 25 Februari 2026

JUDUL DALAM BAHASA INDONESIA :

Optimasi Deteksi Aksen Daerah pada Ucapan Bahasa Indonesia Menggunakan CNN dan Augmentasi Audio

JUDUL DALAM BAHASA INGGRIS :

Optimization of Regional Accent Detection in Indonesian Speech Using CNN and Audio Augmentation

TANDA TANGAN DAN TANGGAL

Pembimbing	Ka. Prodi	Mahasiswa
TGL : 2 Maret 2026	TGL : 25 Februari 2026	TGL : 2 Maret 2026
		

ABSTRAK

Indonesia merupakan negara yang kaya akan keberagaman bahasa dan aksen daerah. Deteksi aksen secara otomatis memiliki potensi besar dalam bidang linguistik, pendidikan, dan pengembangan teknologi berbasis suara. Penelitian ini bertujuan untuk mengoptimalkan proses deteksi aksen daerah pada ucapan Bahasa Indonesia dengan memanfaatkan algoritma Convolutional Neural Network (CNN) dan teknik augmentasi audio. Dataset terdiri dari rekaman suara beraksen dari berbagai daerah di Indonesia, yang telah melalui proses pra-pemrosesan dan ekstraksi fitur Mel Frequency Cepstral Coefficients (MFCC). Untuk meningkatkan performa klasifikasi, dilakukan augmentasi data melalui pitch shifting, time-stretching, penambahan noise, dan manipulasi volume.

Model CNN dirancang untuk mengenali pola khas dari masing-masing aksen dan dilatih menggunakan data hasil augmentasi. Evaluasi dilakukan berdasarkan akurasi, presisi, recall, dan F1-score. Hasil menunjukkan bahwa penggunaan augmentasi audio secara signifikan meningkatkan performa model, dengan akurasi tertinggi mencapai lebih dari 90% pada data uji. Temuan ini membuktikan bahwa kombinasi CNN dan augmentasi audio efektif dalam mendeteksi aksen daerah, serta dapat diterapkan untuk pengembangan aplikasi berbasis suara yang lebih inklusif terhadap keragaman aksen di Indonesia.

Kata kunci: Deteksi Aksan, CNN, Augmentasi Audio, MFCC, Ucapan Bahasa Indonesia.

ABSTRACT

Indonesia is a country rich in linguistic and regional accent diversity. Automatic accent detection plays a crucial role in the fields of linguistics, education, and the development of voice-based technologies. This study aims to optimize regional accent detection in Indonesian speech using the Convolutional Neural Network (CNN) algorithm combined with audio augmentation techniques. The dataset comprises speech recordings representing various regional accents in Indonesia, which were preprocessed and transformed into Mel Frequency Cepstral Coefficients (MFCC) features. To enhance classification performance, audio augmentation was applied, including pitch shifting, time-stretching, noise injection, and volume scaling.

A CNN model was designed to capture unique accent patterns and trained using the augmented dataset. The evaluation was conducted using accuracy, precision, recall, and F1-score metrics. The results show that audio augmentation significantly improves model performance, with the highest accuracy exceeding 90% on the test set. These findings demonstrate that the integration of CNN and audio augmentation is effective for accent detection and can be implemented in voice-based applications that aim to support Indonesia's linguistic diversity.

Keywords: Accent Detection, CNN, Audio Augmentation, MFCC, Indonesian Speech.

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa, karena atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul:

“Optimasi Deteksi Aksent Daerah pada Pengucapan Bahasa Indonesia Menggunakan Augmentasi Audio dan Fine-Tuning Few-Shot Learning.”

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Informatika. Penelitian ini bertujuan untuk menganalisis dan membandingkan performa metode *Convolutional Neural Network* (CNN) dengan augmentasi audio dan pendekatan Fine-Tuning Few-Shot Learning dalam mendeteksi aksent daerah pada kondisi dataset terbatas. Hasil penelitian menunjukkan bahwa pendekatan *Fine-Tuning Few-Shot Learning* mampu meningkatkan akurasi secara signifikan dibandingkan CNN konvensional.

Dalam proses penyusunan skripsi ini, penulis menyadari bahwa banyak pihak yang telah memberikan bantuan, bimbingan, serta dukungan baik secara langsung maupun tidak langsung. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih kepada:

1. Tuhan Yang Maha Esa atas segala rahmat dan kemudahan yang diberikan.
2. Kedua orang tua yaitu bapak Solihun dan ibu Aam Azizah yang selalu memberikan doa, dukungan, dan motivasi. Serta maaf jika saat pengerjaan skripsi banyak marah – marah dan bentak kepada kedua orang tua .
3. Kepada kakak penulis yaitu Eko Panji Setiawan yang selalu mendukung penulis melalui material yang diberikan.

4. Kepada tante penulis yang sudah membantu doa serta dukungan tanpa putus dan juga membantu semangat penulis dengan sikapnya.
5. Terimakasih kepada penulis skripsi sendiri, yaitu Dwi Tunjung Nawang Wulan tidak apa – apa sempat gagal dulu namun saat ini sudah menginjak moment mengerjakan skripsi hingga selesai salah satu bersyukur dan *whislist* penulis nomor satu di tahun 2026, terimakasih atas semangat dan usahanya walaupun



banyak perjuangan dijalani dengan kesedihan tetapi harus tetap bersyukur dan tidak pantang menyerah untuk menyelesaikannya.

6. Dosen pembimbing yang telah memberikan arahan, bimbingan, serta masukan yang sangat berharga selama proses penelitian.
7. Seluruh dosen dan civitas akademika Program Studi Informatika yang telah memberikan ilmu dan wawasan selama masa perkuliahan.
8. Kepada saudara perempuan penulis, yaitu Haura Giandra Aluni yang masih berusia 4 tahun telah menghibur dan membuat penulis bahagia atas kehadirannya tawa dan canda nya sangat dibutuhkan disaat penulis sedang jelek moodnya.
9. Kepada Nabila Lintang Anzani penulis mengucapkan terimakasih karna selama penulisan sudah banyak membantu dari mengumpulkan dataset, processing jika terkendala eror, hingga penulisan, dan dukungan doa yang diberikan.
10. Kepada teman penulis yang berteman dengan penulis lebih dari 10 tahun yaitu Anandita Putri Anjali, yang sudah mau membantu saya selama penulisan bab 1 hingga terus mendukung suka duka penulis selama pengerjaan skripsi.
11. Kepada teman satu *ekstrakurikuler* taekwondo serta ketua osis semasa bangku sekolah menengah pertama, yaitu Meilda Putri Budi Ningsih, yang sudah mau diganggu atas waktunya untuk menemani penulis saat sedang letih dan buntu pemikiran diajaknya penulis keluar rumah.
12. Terimakasih kepada kekasih penulis, yaitu Muhammad Arsy yang sudah membagi waktunya untuk menemani penulis mengerjakan skripsi dari awal pembuatan hingga saat ini penulisan skripsi.
13. Terimakasih kepada boyband BTS dan Enhypen yang menemani penulis saat pengerjaan skripsi sambil mendengarkan karya boyband tersebut.

14. Kepada Kim Taehyung dan Jake Shim penulis mengucapkan terimakasih karena sudah membanggakan penulis walaupun dengan paras kalian penulis sudah sangat bangga terhadap kalian.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna, oleh karena itu kritik dan saran yang membangun sangat penulis harapkan demi perbaikan di



masa yang akan datang. Semoga skripsi ini dapat memberikan manfaat bagi penulis, mahasiswa lain, maupun pihak – pihak yang berkepentingan.

Bogor, 25 Februari 2026

Dwi Tunjung Nawang Wulan



HALAMAN JUDUL	i
ABSTRAK	iii
ABSTRACT	iv
KATA PENGANTAR	v
DAFTAR ISI	vii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	x
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah.....	4
1.3 Rumusan Masalah.....	5
1.4 Tujuan Masalah	6
1.5 Batasan Penelitian.....	6
1.6 Manfaat Penelitian.....	7
BAB II TINJAU PUSTAKA.....	8
2.1 Penelitian Sebelumnya.....	20
2.2 Studi Literatur	21
2.3 Machine Learning	31
2.4 Convolution Neural Network (CNN).....	31
2.3.1 Lapisan Konvolusi	32
2.3.2 Lapisan Polling	32
2.3.3 Lapisan Aktivasi	32
2.3.4 Lapisan Fully	32
2.5 Pendekatan Few-Shot Learning Sebagai Pemrosesan Audio	34
2.6 Dukungan Teknologi Untuk Penerapan Machine Learning yang Efektif dan Efisien.....	34
2.4.1 Python Sebagai Bahasa Pemrograman Utama.....	34
2.4.2 Google Colaboratory Untuk Komputasi Berkinerja Tinggi....	35
.....	35

2.7	Augmentasi Audio	35
BAB III METODE PENELITIAN.....		37
3.1	Timeline Penelitian	38
3.2	Sumber Data.....	38
3.3	Pendekatan Penelitian	38
3.4	Penentuan Objek	39
3.5	Populasi dan Sampel	39
3.6	Diagram Penelitian.....	40
3.6.1	Studi Pustaka dan Pencarian Dataset Eksisting (Skunder)	42
3.6.2	Perekaman Data Primer	42
3.6.3	Evaluasi Dataset (Aksen, Jumlah Penutur, Kualitas Audio dan Lisensi Audio).....	42
3.6.4	Pelabelan dan Validasi Metadata Penutur	44
3.7	Pra-pemrosesan	44
3.7.1	Augmentasi Audio	44
3.7.5	Ekstrasi Fitur.....	45
3.7.6	Resampling Episodik	46
3.7.7	Input ke Prototypical Network.....	46
3.7.8	Embedding Network	47
3.7.9	Pembagian Dataset.....	47
3.8	Training Data	47
3.8.1	Input CNN.....	47
3.8.5	<i>Input Model Few-Shot Learning</i>	47
3.9	Evaluasi Model	48
3.9.1	Akurasi.....	48
3.9.2	Presisi	49
3.10	Matrix Konfusi (Confusion Matrix).....	49
3.11	Implementasi	49
3.12	Diagram Fishbone	50
3.11.1	Man (Data).....	50
3.11.2	Methods (Algoritma)	50

3.11.3	Machine (Komputasi)	50
3.11.4	Man (Sumber Daya)	51
3.11.5	Measurement (Evaluasi)	51
3.11.6	Mother Nature (Lingkungan/Konteks)	51
BAB IV HASIL DAN PEMBAHSAN		53
4.1	Dataset	53
4.1.1	Persiapan Dataset	53
4.1.2	<i>Preprocessing Audio</i>	55
4.1.3	<i>Process Ekstrasi Fitur</i>	59
4.1.4	Augmentasi Audio	63
4.2	Processing Model dengan Algoritma CNN	66
4.3	Perhitungan Manual Pada Model CNN	69
4.3.1	Perhitungan Total Parameter Model	70
4.3.2	Verifikasi Dimensi Output Conv2D & Pooling	70
4.4	Pengujian Model Kedua <i>Few-Shot Learning</i>	70
4.4.1	<i>Preprocessing Few-Shot Dengan Episodic</i>	71
4.4.2	Analisis Jumlah Karna	75
4.4.3	Analisis Hasil Perhitungan <i>Few-Shot Learning</i> Per Episodic	77
4.4.4	Pembuatan Embedding Network	78
4.4.1	Korelasi Arsitektur CNN dengan Ekstrasi Fitur	79
4.4.6	Pembuatan Prototypical Network	80
4.4	Pembagian Data	81
4.5	Processing Training Model <i>Few-Shot Learning</i> dan Augmentasi Audio	81
4.5.1	Input Model <i>Few-Shot</i>	82
4.5.2	History Training <i>Few-Shot Learning</i>	85
4.6	Evaluasi Kinerja Model	86
4.7	Analisis Confusion Matrix	87
4.8.1	Interpretasi Matriks	88
4.8	Perhitungan Confusion Matrix	89
4.9	Visual TSNE	89

4.10	Evaluasi Perbandingan Model CNN dan <i>Fine-Tuning Few-Shot Learning</i>	91
4.10.1	Keunggulan Strategi Few-Shot Learning	91
4.10.2	Perbandingan Augmentasi Audio Model CNN dengan Few-Shot	92
4.10.3	Evaluasi Kompleksitas dan Efisiensi	93
4.10.4	Kesimpulan Evaluasi	96
4.11	Implementasi Deteksi Aksent	96
BAB V KESIMPULAN DAN SARAN		103
5.1	Kesimpulan	103
5.2	Saran	104
DAFTAR PUSTAKA		105
LAMPIRAN		108



DAFTAR GAMBAR

<i>Gambar 3.1</i> Diagram Penelitian.....	41
<i>Gambar 3.2</i> Diagram Fishbone.....	49
<i>Gambar 4. 1</i> Persiapan Dataset	52
<i>Gambar 4. 2</i> Labelling Dataset	53
<i>Gambar 4. 3</i> Distribusi Kelas Aksen	54
<i>Gambar 4. 4</i> Fitur MFCC dengan Lima Kelas Aksen.....	55
<i>Gambar 4. 5</i> Source Code Ekstrasi Fitur Audio CNN	60
<i>Gambar 4. 6</i> Source Code Target Multitask CNN	61
<i>Gambar 4. 7</i> Source Code Augmentasi Audio	62
<i>Gambar 4. 8</i> Variasi Augmentasi Audio	64
<i>Gambar 4. 9</i> Source Code Input Model CNN	65
<i>Gambar 4. 10</i> Input Training Model CNN	66
<i>Gambar 4. 11</i> Result Evaluasi Model CNN	67
<i>Gambar 4. 12</i> Source Code Ekstrasi Fitur Few-Shot.....	70
<i>Gambar 4. 13</i> Source Code Target Multitask Few-Shott	73
<i>Gambar 4. 14</i> Mel-spectrogram Pada Audio Jakarta(1)	76
<i>Gambar 4. 15</i> Membuat Episodic	77
<i>Gambar 4. 16</i> Source Code Pembuatan Embedding Network	78
<i>Gambar 4. 17</i> Source Code Pembuatan Prototypical Network	79
<i>Gambar 4. 18</i> Source Code Split Data	80
<i>Gambar 4. 19</i> Input Model Few-Shot	81
<i>Gambar 4. 20</i> Input Model Few-Shot	82
<i>Gambar 4. 21</i> Input Model Few-Shot	82
<i>Gambar 4. 22</i> Input Model Few-Shot	82
<i>Gambar 4. 23</i> Source Code Hasil Training Few-Shot	83
<i>Gambar 4. 24</i> History Training Few-Shot Learning	84
<i>Gambar 4. 25</i> History Training Few-Shot Learning kedua	85
<i>Gambar 4. 26</i> Distribusi Akurasi 30 Episodic	86

<i>Gambar 4. 27</i>	<i>Confusion Matrix</i>	87
<i>Gambar 4. 28</i>	<i>Visual TSNE</i>	89
<i>Gambar 4. 29</i>	<i>Val dan Train Loss Model CNN</i>	93
<i>Gambar 4. 30</i>	<i>Val dan Train Loss Model CNN</i>	93
<i>Gambar 4. 31</i>	<i>Val dan Train Loss Model Few-Shot</i>	93
<i>Gambar 4. 32</i>	<i>Audio Input</i>	96
<i>Gambar 4. 33</i>	<i>Deteksi Aksen Sekarang</i>	96
<i>Gambar 4. 34</i>	<i>Hasil Deteksi Aksen Jakarta</i>	97
<i>Gambar 4. 35</i>	<i>Hasil Deteksi Aksen Jawa Barat</i>	98
<i>Gambar 4. 36</i>	<i>Hasil Deteksi Aksen Jawa Tengah</i>	99
<i>Gambar 4. 37</i>	<i>Hasil Deteksi Aksen Jawa Timur</i>	100
<i>Gambar 4. 38</i>	<i>Hasil Deteksi Aksen D.I YogyaKarta</i>	101



<i>Gambar 4. 25</i>	<i>History Training Few-Shot Learning kedua</i>	85
<i>Gambar 4. 26</i>	<i>Distribusi Akurasi 30 Episodic</i>	86
<i>Gambar 4. 27</i>	<i>Confusion Matrix</i>	87
<i>Gambar 4. 28</i>	<i>Visual TSNE</i>	89
<i>Gambar 4. 29</i>	<i>Val dan Train Loss Model CNN</i>	93
<i>Gambar 4. 30</i>	<i>Val dan Train Loss Model CNN</i>	93
<i>Gambar 4. 31</i>	<i>Val dan Train Loss Model Few-Shot</i>	93
<i>Gambar 4. 32</i>	<i>Audio Input</i>	96
<i>Gambar 4. 33</i>	<i>Deteksi Aksen Sekarang</i>	96
<i>Gambar 4. 34</i>	<i>Hasil Deteksi Aksen Jakarta</i>	97
<i>Gambar 4. 35</i>	<i>Hasil Deteksi Aksen Jawa Barat</i>	98
<i>Gambar 4. 36</i>	<i>Hasil Deteksi Aksen Jawa Tengah</i>	99
<i>Gambar 4. 37</i>	<i>Hasil Deteksi Aksen Jawa Timur</i>	100
<i>Gambar 4. 38</i>	<i>Hasil Deteksi Aksen D.I YogyaKarta</i>	101



DAFTAR TABEL

Tabel 2.2 Studi Literatur.....	21
Tabel 3.1 Timeline Penelitian.....	36
Tabel 4. 1 Delta - delta MFCC Pattern (Sunda)	56
Tabel 4. 2 Delta - delta MFCC Pattern (YogyaKarta).....	58
Tabel 4. 3 Result Accuracy Model CNN.....	69
Tabel 4. 4 Struktur Per-episodic	84
Tabel 4. 5 Interpretasi Matriks	88
Tabel 4. 6 Perbandingan Model CNN dan Few-Shot.....	91
Tabel 4. 7 Evaluasi Kompleksitas dan Efisiensi	93

