
Sistem Pengenalan Suara menggunakan Metode Mel Frequency Cepstral Coefficients dan Gaussian Mixture Model

Ababil Azies Sasilo¹, Rizal Adi Saputra^{2*}, Ika Purwanti Ningrum³

^{1,2,3}Jurusan Teknik Informatika, Fakultas Teknik, Universitas Halu Oleo
Jl. HEA Mokodompit No. 8, Kendari, Indonesia 93232

*email: rizaladisaputra@uho.ac.id

(Naskah masuk: 23 Februari 2022; diterima untuk diterbitkan: 06 Juli 2022)

ABSTRAK – Teknologi biometrik sedang menjadi tren teknologi dalam berbagai bidang kehidupan. Teknologi biometrik memanfaatkan bagian tubuh manusia sebagai alat ukur sistem yang memiliki keunikan disetiap individu. Sistem pengenalan suara adalah salah satu penerapan teknologi biometrik yang fokus kepada suara manusia. Sistem pengenalan suara menjadi teknologi biometrik yang memiliki potensi besar dan memerlukan metode ekstraksi fitur dan metode klasifikasi, salah satu metode ekstraksi fitur adalah metode Mel Frequency Cepstral Coefficients (MFCC). Metode MFCC merupakan metode ekstraksi fitur. Metode ini dimulai dari tahap pre-emphasis, frame blocking, windowing, fast fourier transform, mel frequency wrapping dan cepstrum. Sedangkan metode klasifikasi menggunakan Gaussian Mixture Model (GMM) dengan menghitung likelihood kesamaan antar data ekstraksi ciri suara. Tujuan penelitian ini untuk mengetahui seberapa baik metode MFCC-GMM dalam sistem pengenalan suara. Pengujian sistem dibagi menjadi dua yaitu pengujian pada kondisi ideal dan pada kondisi tidak ideal. Kondisi ideal adalah kondisi dimana tidak terdapat gangguan noise maupun derau saat melakukan pengujian, sedangkan kondisi tidak ideal adalah kondisi dimana terdapat gangguan noise maupun derau saat proses pengujian. Berdasarkan hasil pengujian, metode MFCC - GMM pada kondisi ideal memiliki tingkat akurasi sebesar 82.22% sedangkan pada kondisi tidak ideal mendapatkan akurasi sebesar 66.67%.

Kata Kunci – Mel Frequency Cepstral Coefficients, Gaussian Mixture Model, Pengenalan suara.

Voice Recognition System using Mel Frequency Cepstral Coefficients Method and Gaussian Mixture Model

ABSTRACT - Biometric technology is becoming a technological trend in various fields of life. Biometric technology utilizes human body parts as a system measurement tool that is unique to each individual. Voice recognition system is one of the applications of biometric technology that focuses on the human voice. Voice recognition systems are a biometric technology that has great potential and requires feature extraction methods and classification methods, one of the feature extraction methods is the Mel Frequency Cepstral Coefficients (MFCC) method. The MFCC method is a feature extraction method. This method starts from the pre-emphasis, frame blocking, windowing, fast fourier transform, mel frequency wrapping and cepstrum stages. While the classification method uses the Gaussian Mixture Model (GMM) by calculating the similarity likelihood between sound feature extraction data. The purpose of this research is to find out how good the MFCC-GMM method is in the voice recognition system. System testing is divided into two, namely testing under ideal conditions and under non-ideal conditions. Ideal conditions are conditions where there is no noise or noise disturbance when testing, while non-ideal conditions are conditions where there is noise or noise disturbance during the testing process. Based on the test results, the MFCC - GMM method in ideal conditions has an accuracy rate of 82.22% while in non-ideal conditions it gets an accuracy of 66.67%.

Keywords – Gaussian Mixture Model, Mel Frequency Cepstral Coefficients, Voice Recognition.

1. PENDAHULUAN

Dewasa ini, perkembangan teknologi informasi menjadi sangat pesat dan masif. Banyak peralatan teknologi saat ini mengusung ukuran yang berbasis pada tubuh manusia yang sering disebut dengan biometrik. Teknologi biometrik memanfaatkan bagian tubuh manusia seperti wajah, sidik jari, retina mata dan lain-lain digunakan untuk berbagai keperluan seperti keamanan, kesehatan, maupun pendidikan[1]. Suara merupakan salah satu ukuran biometrik yang belum mendapat perhatian lebih dalam teknologi biometrik. Sinyal suara merupakan sebuah sinyal yang dapat terdengar dan dapat diukur menggunakan satuan waktu dan/atau ruang[2].

Selama masa pandemi, kebutuhan akan teknologi nirsentuh semakin meningkat demi menurunkan angka penyebaran Covid-19. Berdasarkan data dari *Capgemini Research Institute* pada tahun 2020, ekspektasi akan kebutuhan teknologi nirsentuh skala global meningkat hingga 77% disaat *pandemic* dan 62% pasca pandemik, selain itu berdasarkan data yang sama, 59% masyarakat ingin menggunakan aplikasi berbasis suara ketika mengakses fasilitas publik untuk menghindari kontak fisik dan interaksi sesama manusia[3].

Pengenalan suara atau yang dikenal dengan *Automatic Speech Recognition* adalah pengembangan teknik dan sistem pembelajaran yang diterapkan pada mesin atau komputer untuk dapat menerima masukan berupa suara yang disusun menjadi kata maupun kalimat yang diucapkan[4]. Teknologi ini memungkinkan sebuah mesin dapat belajar dan mengenali sumber suara dengan cara mengolah suara yang berupa kata atau kalimat menjadi sinyal suara dan, lalu sinyal suara tersebut akan diubah kedalam bentuk digital untuk kemudian dilakukan pencocokan sinyal tersebut dengan suatu pola tertentu yang telah tersimpan dalam mesin pembelajaran.

Berdasarkan penelitian Purwono Prasetyawan menggunakan metode MFCC dan *Subband Based Cepstral* (SBC) dengan ciri pencocokan menggunakan *Linde Buzo Gray-Vector Quantization* (LBG-VQ)[5]. Penelitian tersebut menyimpulkan bahwa MFCC memiliki akurasi dan kecepatan proses identifikasi yang lebih baik dibandingkan SBC dengan akurasi yang didapatkan oleh MFCC masih dibawah 70%.

M. Sardjono dkk menerbitkan laporan penelitian yang berjudul Pengenalan Suara Pembicara Menggunakan Ekstraksi MFCC dengan Metode GMM pada tahun 2017[6]. Hasil Penelitian menjelaskan bahwa tingkat akurasi pengenalan suara menggunakan MFCC-GMM sebesar 87.75%.

Dari penjelasan di atas, penulis akan membangun "Sistem Pengenalan Suara menggunakan Metode

Mel Frequency Cepstral Coefficients dan *Gaussian Mixture Model*". Metode MFCC akan digunakan sebagai metode ekstraksi ciri suara yang memiliki akurasi tinggi dan Metode GMM akan digunakan sebagai metode klasifikasi yang fleksibel terhadap berbagai jenis data dan kelas. MFCC menjadi tahapan awal dan GMM menjadi tahapan akhir sesuai dengan alur kerja sistem.

2. METODE DAN BAHAN

Suara adalah suatu gelombang yang dihasilkan oleh sebuah benda yang bergetar. Suara dapat merambat dan membutuhkan medium untuk merambat, salah satu medium perambatan suara adalah udara. Suara tidak bisa merambat melalui ruang hampa. Gelombang suara bervariasi dalam tingkatan tekanan suara (amplitudo) dan dalam frekuensi. Jumlah waktu yang diperlukan untuk terjadinya suatu getaran atau gelombang dinamakan periode (T). Sedangkan jumlah gelombang yang terjadi setiap detik dinamakan frekuensi (f) dengan satuan m/dt (Hz)[7].

Sebuah sinyal yang rumit dihasilkan oleh individu dihasilkan dari beberapa transformasi pada beberapa level yang berbeda dari semantik, linguistik, artikulasi dan akustik. Perbedaan ini terjadi akibat perbedaan masing-masing pita suara dan kebiasaan pengucapan yang dipelajari secara berbeda di masing-masing individu. Suara dapat diukur berdasarkan tingkat intensitas suara (*loudness*) dengan menggunakan satuan desibel (dB). Suara manusia saat melakukan percakapan normal umumnya berada pada rentang 50-65 desibel[8].

Audio dapat diartikan sebagai suara atau hasil reproduksi suara. Gelombang suara adalah gelombang yang dihasilkan dari sebuah benda yang bergetar. Gelombang suara dapat merambat pada medium udara, air, atau mineral lainnya kecuali ruangan hampa udara yang merupakan ruangan tanpa tekanan dan medium. Gelombang suara memiliki lembah dan bukit. Satu buah lembah dan bukit akan menghasilkan satu siklus atau periode. Siklus ini akan berlangsung secara terus menerus sehingga membentuk konsep frekuensi. Frekuensi adalah jumlah siklus yang terjadi dalam satu detik dan memiliki satuan Hertz atau disingkat Hz.[7]

Manusia memiliki alat pendengaran yaitu telinga. Telinga manusia dapat mendengar bunyi antara 20 Hz sampai 20 KHz sesuai dengan limitasi sinyal suara yang dapat diterima oleh gendang telinga manusia. Suara dibawah 20 Hz seringkali dihasilkan oleh serangga dan suara diatas 20 KHz dihasilkan oleh paus atau ikan duyung[8].

Pengenalan suara merupakan salah satu upaya cara agar suara dapat dikenali atau diidentifikasi sehingga dapat dimanfaatkan. Pengenalan suara dapat dibedakan kedalam tiga bentuk pendekatan,

yaitu pendekatan akustik-fonetik, pendekatan kecerdasan buatan, dan pendekatan pengenalan pola[9].

Pendekatan pengenalan pola terdiri dari dua tahap yaitu tahap pembelajaran pola dan pengenalan pola melalui perbandingan pola. Tahap perbandingan pola adalah tahap bagi suara untuk dikenali lalu dibandingkan polanya dengan setiap kemungkinan pola yang telah didapatkan pada tahap pembelajaran, untuk kemudian diklasifikasikan dengan pola terbaik yang cocok[10].

Dalam pengenalan suara mempunyai beberapa metode yang dapat digunakan dalam tahap ekstraksi ciri, diantaranya adalah *Mel Frequency Cepstral Coefficients* (MFCC). MFCC merupakan metode yang merepresentasikan audio sebagai sebuah koefisien yang didasarkan pada frekuensi suara yang dapat didengar oleh manusia. MFCC merupakan salah satu metode terbaik dalam ekstraksi ciri suara. Metode ini diperkenalkan oleh Davis dan Mermelstein di tahun 1980-an. Ekstraksi ciri dalam metode ini ditandai dengan perubahan data suara menjadi data berupa spectrum gelombang[8].

Alur pemrosesan MFCC dibuat mengadopsi alur pemrosesan sistem indra pendengaran manusia dalam menangkap sinyal suara agar hasil ekstraksi fiturnya mendekati persepsi yang dihasilkan oleh indra manusia. MFCC mampu menghasilkan data seminimal mungkin tanpa menghilangkan informasi-informasi penting yang ada pada sinyal suara dan hal ini menjadi salah satu kelebihan MFCC dibanding metode lain[9].

Data

Data yang digunakan dalam penelitian ini merupakan data rekaman suara yang diambil dari mahasiswa Jurusan Teknik Informatika, Universitas Halu Oleo. Data yang digunakan memiliki format "wav" dan berdurasi selama 5 detik. Data rekaman suara memiliki rentang intensitas suara diantara 63-78 dB. Data diambil dari 15 orang dan masing-masing orang merekam 27 suara dengan kalimat "Nomor Induk Mahasiswa, Nama Mahasiswa".

Mel Frequency Cepstral Coefficients (MFCC)

Alur Pemrosesan MFCC terdiri dari *pre-emphasis*, *frame blocking*, *windowing*, *fast fourier transform*, *mel frequency wrapping* dan terakhir *cepstrum*[6].

Metode ini akan menghasilkan data ekstraksi ciri suara berupa vektor mel atau vektor ciri suara yang nantinya akan digunakan sebagai data input GMM untuk membuat model suara[11].

Gaussian Mixture Models (GMM)

Gaussian Mixture Model (GMM) merupakan metode estimasi jarak kedekatan dan salah satu yang secara umum digunakan sebagai pemodelan atau

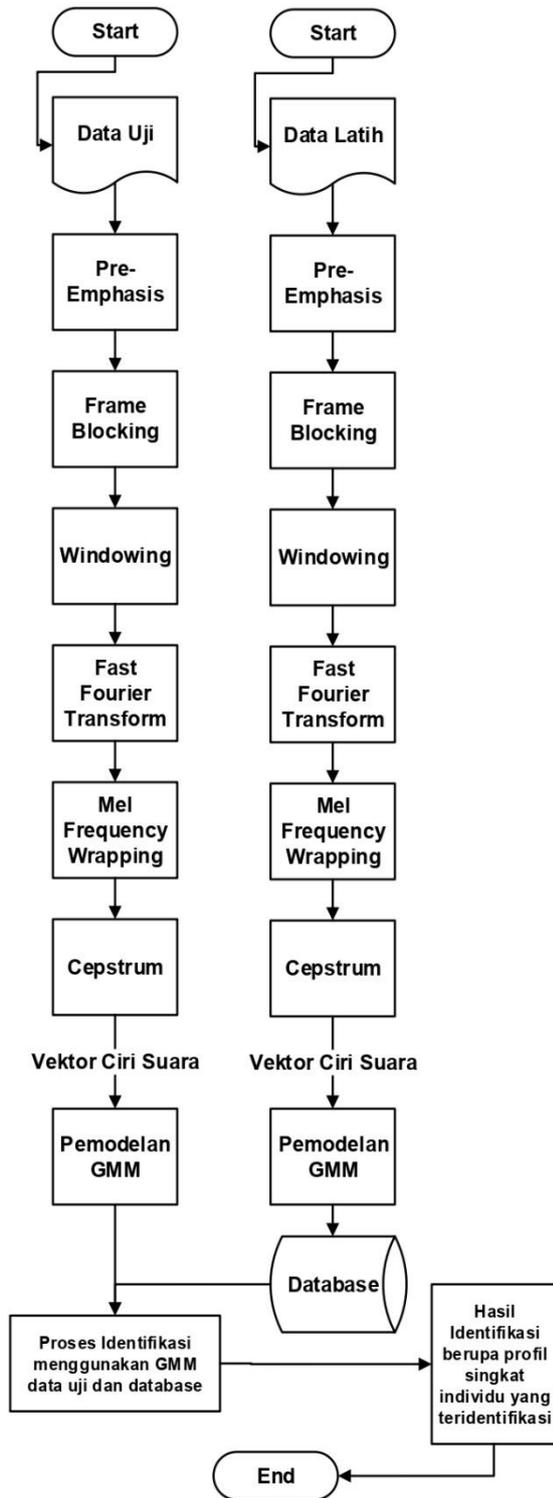
klasifikasi. GMM memanfaatkan estimasi *Maximum likelihood* sebagai parameter[6].

Agar mendapatkan model dari setiap suara pembicara secara optimal maka peneliti memerlukan perhitungan estimasi yang baik dari parameter GMM. Untuk menghasilkan hal tersebut, diperlukan metode yang efisien yaitu pendekatan *Maximum Likelihood Estimation* (ML). Pada identifikasi pembicara, setiap pembicara diwakili oleh sebuah GMM dan hal itu berdasarkan pada model suara pembicara tersebut[10].

Data suara pembicara yang telah diekstraksi menggunakan MFCC berupa vektor ciri suara, akan diproses oleh GMM dan kemudian menghasilkan model data latih suara pembicara sehingga dapat diestimasi parameter pada GMM, dinotasikan dengan x , yang dalam beberapa hal paling cocok untuk setiap vector fitur data latih. Terdapat beberapa teknik yang menyediakan untuk estimasi parameter dari GMM. Sejauh ini metode yang paling populer dan handal adalah estimasi *maximum likelihood* (ML). Tujuan utama metode ini untuk mencari parameter model yang nilai kemungkinan cocoknya maksimal dengan model GMM dari data latih.

Langkah Analisis

Sistem yang akan dibangun adalah sistem pengenalan suara menggunakan metode MFCC dan GMM. Alur kerja sistem dapat dilihat pada Gambar 1. Pada sistem ini terdapat tiga tahap utama yaitu akuisisi data, ekstraksi ciri, pemodelan ciri, menyimpan ke basisdata, dan pengenalan suara. Pada tahap akuisisi data, data yang digunakan rekaman suara dengan format WAV yang akan direkam oleh sistem secara langsung. Pada tahap ekstraksi ciri, rekaman suara yang telah masuk akan diproses menggunakan metode MFCC yang terdiri dari tahap *pre-emphasis*, *frame blocking*, *windowing*, *fast fourier transform*, *mel frequency wrapping* dan *cepstrum*. Tahapan ini akan menghasilkan vektor mel atau vektor ciri suara yang akan digunakan pada tahap pemodelan GMM. Pada tahap pemodelan, vektor ciri suara hasil dari tahapan akhir MFCC akan diproses menggunakan GMM untuk mendapatkan model dengan format .gmm yang akan tersimpan di basis data sistem dan sistem akan mengecek kemiripan model dari data latih dan data uji dan menemukan data yang memiliki kemiripan paling tinggi, lalu sistem akan memberikan hasil berupa tingkat kemiripan serta profil singkat data yang berhasil dikenali[12].



Gambar 1. Ilustrasi Langkah Analisis Sistem

3. HASIL DAN PEMBAHASAN

Akuisisi Data

Akuisisi data dilakukan di laboratorium artificial intelligence Jurusan Teknik Informatika, Universitas Halu Oleo. Jumlah data rekaman suara sebanyak 33 rekaman suara dengan durasi 5 detik dan berasal dari masing-masing 15 orang mahasiswa jurusan teknik informatika universitas halu oleo dengan kalimat rekaman suara "Nama Mahasiswa,

Nomor Induk Mahasiswa". Dari 33 data rekaman suara, akan dibagi menjadi data latih dan data uji dengan jumlah 27 rekaman sebagai data latih, 3 rekaman sebagai data uji dengan kondisi ideal dan 3 rekaman sebagai data uji dengan kondisi tidak ideal. Data uji dan data latih akan dimasukkan kedalam basisdata sistem yang terbuat dari bahasa pemrograman python[13]

Pre-Processing Data

Pada data latih, diterapkan filter *pre-emphasis* untuk mengurangi derau pada data rekaman suara dan menyeimbangkan frekuensi rendah dan tinggi dari data latih.

Ekstraksi Ciri Dengan MFCC

Pada proses ini dilakukan pengambilan fitur pada data latih menggunakan MFCC. Proses ini terdiri beberapa tahapan yaitu *pre-emphasis*, *frame blocking*, *windowing*, *fast fourier transform*, *mel frequency wrapping* dan *cepstrum*[4]. *Pre-emphasis* adalah tahapan preprocessing yang berfungsi sebagai filter yang dapat menghilangkan derau pada suara rekaman dan menyeimbangkan frekuensi tinggi dan rendah pada suara manusia. *Frame blocking* bertujuan untuk membagi sinyal suara yang telah melalui tahapan *pre-emphasis* menjadi beberapa frame dengan jarak waktu 25 milidetik dan jarak antar frame adalah 10 milidetik. *Windowing* bertujuan untuk mengurangi diskontinuitas sinyal pada permulaan dan akhir setiap frame. Misalkan sebuah file audio telah memiliki 348 frame maka setiap frame akan melalui tahap *windowing*. FFT berfungsi untuk mengubah setiap frame dari domain waktu menjadi domain frekuensi. FFT adalah sebuah algoritma cepat untuk implementasi *Discrete Fourier Transform* (DFT) yang dioperasikan pada sinyal waktu diskret yang terdiri dari N sampel. Size FFT yang digunakan pada penelitian adalah 512, hasil dari *windowing* akan melalui proses FFT dan akan memiliki ukuran vektor 512×348 dari yang awalnya 1102×348 . *Mel frequency wrapping* bertujuan untuk mengaplikasikan filter bank yang digunakan untuk mengekstrak kekuatan dari masing-masing band frekuensi adapun jumlah filter yang digunakan adalah 26. *Cepstrum* merupakan tahapan terakhir pada MFCC dan bertujuan untuk menormalisasi hasil dari *mel frequency wrapping* menggunakan algoritma DCT untuk menyederhanakan filter bank karena tidak semua filter pada filter bank tidak memiliki kontribusi pada keunikan suara.

Klasifikasi dengan GMM

Langkah-langkah untuk algoritma gaussian mixture model adalah pertama menginisialisasi nilai μ_k , σ_k , π_k secara acak untuk keseluruhan *cluster*, μ adalah mean, σ adalah *variance*, π adalah koefisien

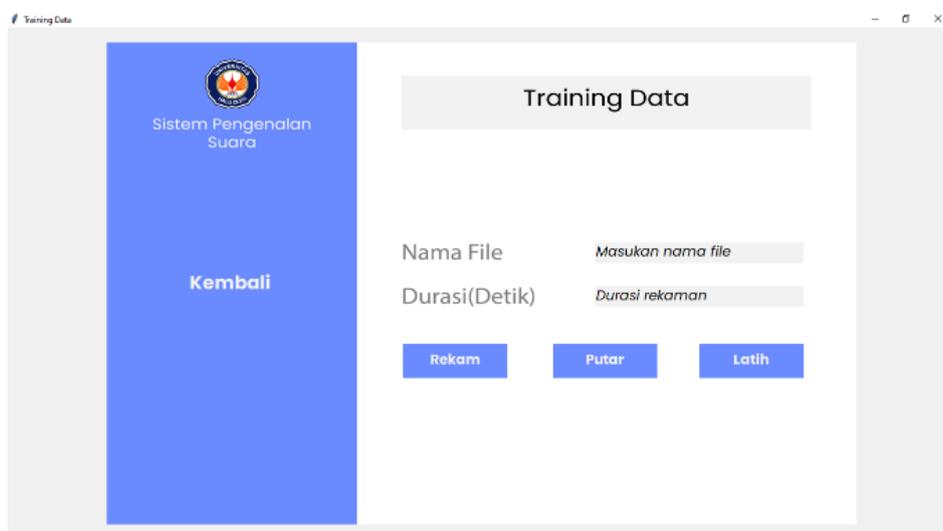
campuran dan k merupakan angka yang menunjuk ke suatu *mixture* dalam distribusi *gaussian* dan k ekuivalen sebagai nilai yang menunjuk suatu cluster. Kedua mengevaluasi hasil *log-likelihood* dengan menggunakan parameter μ_k, σ_k, π_k . C adalah *cluster*, ρ adalah probabilitas, i adalah angka yang menunjuk ke suatu distribusi *gaussian* dan X adalah distribusi *gaussian*. Misalkan cluster C_k diwakili oleh distribusi *gaussian* (μ_k, σ_k) maka probabilitas X_i apapun milik *cluster* C_k dihitung lalu dilanjutkan dengan menghitung nilai *likelihood*. Selanjutnya mengubah nilai $\mu_k, \sigma_k, \rho(C_k)$. Terakhir ulangi langkah 2 dan 3 hingga kriteria konvergensi terpenuhi[14].

Tampilan Aplikasi

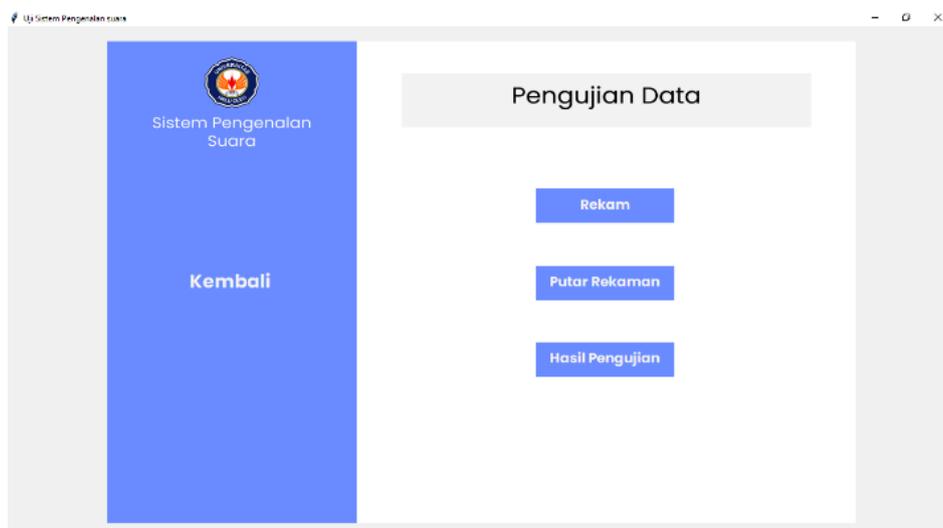
Aplikasi Sistem Pengenalan Suara Dengan Metode MFCC Dan GMM dibangun menggunakan bahasa python dan modul Tkinter sebagai modul tampilan aplikasi. Gambar 2 menampilkan halaman depan aplikasi. Gambar 3 menampilkan halaman menu training data. Gambar 4 menampilkan halaman menu hasil pengujian dan Gambar 5 memperlihatkan tampilan hasil pengujian.



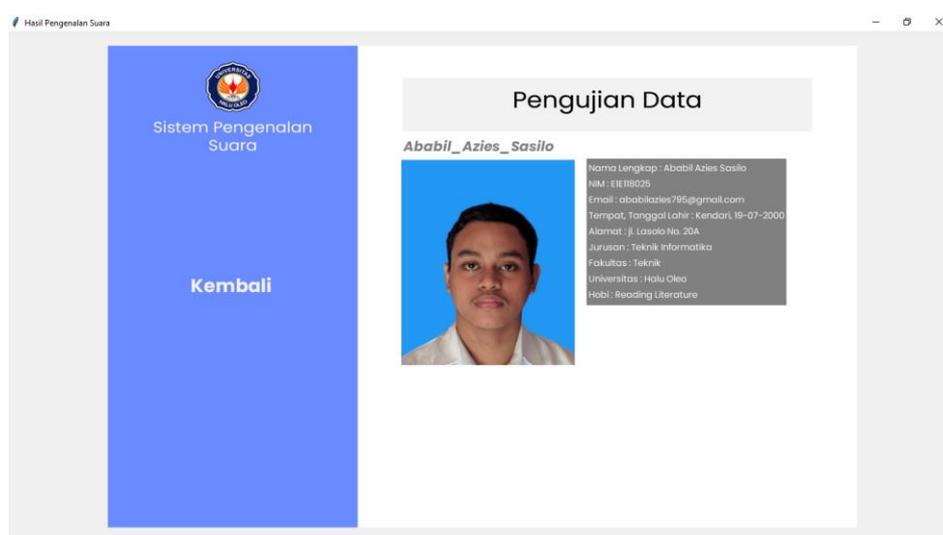
Gambar 2. Halaman Depan Aplikasi



Gambar 3. Menu Training data



Gambar 4. Menu Hasil Pengujuan



Gambar 5. Tampilan Hasil Pengujuan

Tabel 1. Hasil Pengujuan Kondisi Ideal

Hasil Pengujuan

Pengujuan pada penelitian ini dibagi menjadi dua yaitu pengujuan dengan kondisi ideal dan pengujuan dengan kondisi tidak ideal. Pengujuan dengan kondisi ideal adalah pengujuan dengan data uji yang bebas dari derau dan direkam dengan kondisi *noise* rendah (Tabel 1). Pengujuan kondisi tidak ideal adalah pengujuan dengan data uji yang memiliki gangguan berupa suara derau dan kondisi *noise* yang tinggi (Tabel 2). Masing-masing pengujuan menggunakan 3 data uji, sehingga terdapat 45 data uji kondisi ideal dan 45 data uji tidak ideal. Data uji melalui proses yang sama dengan data latih.

No	Nama	dB	Data Uji	Suara Benar	Suara Salah
1	A. G. B. Arifzha	68-71	3	2	1
2	A. A. Sasilo	70-72	3	1	2
3	Asdar	70-73	3	1	2
4	D. E. Putra	69-71	3	3	0
5	H. Octaviani	67-70	3	3	0
6	LD. H. D. Putra	68-72	3	3	0
7	Ld. M., A. Ilham	70-73	3	2	1
8	M. F. Almaliki	69-74	3	3	0
9	M. I. Prayoga	70-74	3	3	0
10	M. I. Salwi	71-74	3	3	0
11	M. Risky.	69-72	3	3	0
12	M. Taslim	73-75	3	3	0
13	N. Madukubah	67-70	3	3	0
14	Reskal	70-72	3	3	0
15	S. N. Rahmah	68-71	3	1	2
Total			45	37	8

Tabel 2. Hasil Pengujian Kondisi Tidak Ideal

No	Nama	dB	Data Uji	Suara Benar	Suara Salah
1	A. G. B. Arifzha	68-71	3	3	0
2	A. A. Sasilo	70-72	3	0	3
3	Asdar	70-73	3	3	0
4	D. E. Putra	69-71	3	3	0
5	H. Octaviani	67-70	3	3	0
6	LD. H. D. Putra	68-72	3	0	3
7	Ld. M.. A. Ilham	70-73	3	3	0
8	M. F. Almaliki	69-74	3	3	0
9	M. I. Prayoga	70-74	3	3	0
10	M. I. Salwi	71-74	3	0	3
11	M. Risky.	69-72	3	0	3
12	M. Taslim	73-75	3	0	3
13	N. Madukubah	67-70	3	3	0
14	Reskal	70-72	3	3	0
15	S. N. Rahmah	68-71	3	3	0
Total			45	30	15

Dari hasil-hasil Tabel 1 dan Tabel 2 dapat disimpulkan persentase keberhasilan menggunakan rumus *receiver operator characteristic* sebagai berikut[15] (Tabel 3):

Persentase Data Uji kondisi Ideal :

$$Precision = (37 / (37+8)) * 100\% = 82.22\%$$

Persentase Data Uji kondisi Ideal :

$$Precision = (30 / (30+15)) * 100\% = 66.67\%$$

Tabel 3. Persentase Pengujian

No	Nama	Persentase	
		Kondisi Ideal	Kondisi Tidak ideal
1	A. G. B. Arifzha	67%	100%
2	A. A. Sasilo	33%	0%
3	Asdar	33%	100%
4	D. E. Putra	100%	100%
5	H. Octaviani	100%	100%
6	LD. H. D. Putra	100%	0%
7	Ld. M.. A. Ilham	67%	100%
8	M. F. Almaliki	100%	100%
9	M. I. Prayoga	100%	100%
10	M. I. Salwi	100%	0%
11	M. Risky.	100%	0%
12	M. Taslim	100%	0%
13	N. Madukubah	100%	100%
14	Reskal	100%	100%
15	S. N. Rahmah	33%	100%
Total		82.22%	66.67%

Pada pengujian dengan kondisi pengambilan data ideal didapatkan persentase tertinggi yaitu 100% dan persentase terendah yaitu 33%. Pada pengujian dengan kondisi pengambilan data tidak ideal didapatkan persentase tertinggi yaitu 100% dan

persentase terendah yaitu 0%.

Tabel 4. Persentase Penelitian Saat ini dan sebelumnya

No	Nama	Tahun	Persentase
1	Prasetyawan (MFCC LBG-VG)[5]	2018	75.71%
2	Maurya et al. (MFCC-GMM)[10]	2017	87.75%
3	Ababil Azies Sasilo et al (MFCC-GMM)	2022	82.22%

Berdasarkan Tabel 4 dapat disimpulkan bahwa penelitian ini memiliki tingkat akurasi yang lebih tinggi dibandingkan penelitian sebelumnya dari Prasetyawan [5] tapi masih di bawah penelitian Maurya et al [10] dengan menggunakan metode dan topik yang serupa.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai Implementasi Metode MFCC Dan GMM Pada Sistem Pengenalan Suara, maka diperoleh kesimpulan sebagai berikut:

Pada kondisi pengambilan data uji ideal didapatkan persentase keberhasilan sistem dalam mengenali suara sebesar 82.22%. pada kondisi pengambilan data uji tidak ideal didapatkan persentase keberhasilan sistem dalam mengenali suara sebesar 66.67%. Sehingga dapat disimpulkan bahwa penggunaan metode MFCC-GMM pada sistem pengenalan suara memiliki tingkat pengenalan suara yang cukup baik untuk studi kasus data berupa identitas mahasiswa.

DAFTAR PUSTAKA

- [1] M. Hernandez-de-Menendez, R. Morales-Menendez, C. A. Escobar, and J. Arinez, "Biometric applications in education," *Int. J. Interact. Des. Manuf.*, vol. 15, no. 2-3, pp. 365-380, 2021, doi: 10.1007/s12008-021-00760-6.
- [2] H. Purwins, B. Li, T. Virtanen, J. Schlüter, S. Y. Chang, and T. Sainath, "Deep Learning for Audio Signal Processing," *IEEE J. Sel. Top. Signal Process.*, vol. 13, no. 2, pp. 206-219, 2019, doi: 10.1109/JSTSP.2019.2908700.
- [3] Capgemini, "COVID-19 and the age of the contactless customer experience," *Capgemini*, 2022. <https://www.capgemini.com/research/covid-19-and-the-age-of-the-contactless-customer-experience/>.
- [4] A. Z. Ridho, "Coefficient Dan Dynamic Time Warping K Nearest Neighbour Dalam Rekognisi Aksen Suku Di Indonesia Skripsi Oleh : Asrori

- Zainur Ridho," Universitas Islam Negeri Maulana Malik Ibrahim Malang, 2019.
- [5] P. Prasetyawan, "Perbandingan Identifikasi Pembicara Menggunakan MFCC dan SBC Dengan Ciri Pencocokan Lbg-Vq," *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2016, no. Sentika, pp. 18–19, 2018, doi: 10.31227/osf.io/85k9u.
- [6] D. K. Putra, I. I. Triasmoro, and R. D. Atmaja, "Simulasi Dan Analisis Speaker Recognition Menggunakan Metode Mel Frequency Cepstrum Coefficient (MFCC) Dan Gaussian Mixture Model (GMM)," *eProceedings Eng.*, vol. 4, no. 2, pp. 1766–1772, 2017, [Online]. Available: <http://libraryproceeding.telkomuniversity.ac.id/index.php/engineering/article/view/487/460>.
- [7] K. Su, X. Liu, and E. Shlizerman, "How Does it Sound? Generation of Rhythmic Soundtracks for Human Movement Videos," *Conf. Neural Inf. Process. Syst.*, vol. 35, no. NeurIPS, pp. 0–10, 2021, [Online]. Available: <https://openreview.net/forum?id=JuNatTaGZ6J>.
- [8] M. I. Marrufo-Pérez, D. del Pilar Sturla-Carretero, A. Eustaquio-Martín, and E. A. Lopez-Poveda, "Adaptation to noise in human speech recognition depends on noise-level statistics and fast dynamic-range compression," *J. Neurosci.*, vol. 40, no. 34, pp. 6613–6623, 2020, doi: 10.1523/JNEUROSCI.0469-20.2020.
- [9] F. D. Adhinata, D. P. Rakhmadani, and A. J. T. Segara, "Pengenalan Jenis Kelamin Manusia Berbasis Suara Menggunakan MFCC dan GMM," *Jorunal data Sci. IoT, Mach. Learn. Informatics*, vol. 1, no. 1, pp. 1–12, 2021.
- [10] A. Maurya, D. Kumar, and R. K. Agarwal, "Speaker Recognition for Hindi Speech Signal using MFCC-GMM Approach," *Procedia Comput. Sci.*, vol. 125, pp. 880–887, 2018, doi: 10.1016/j.procs.2017.12.112.
- [11] M. G. Alenra, "Penerapan Mel Frequency Cepstrum Coefficients (Mfcc) Dan Backpropagation Neural Network (Bpnn) Untuk Pengenalan Huruf Hijaiyah," Universitas Islam Negeri Sultan Syarif Kasim, 2018.
- [12] T. A. P. Nurarinda, J. Sahertian, and U. Mahdiyah, "Rancangan Sistem Identifikasi Jenis Burung Kicau Berdasarkan Suara Burung dengan Mel Frequency Cepstrum Coefficiens (MFCC)," *Semin. Nas. Inov. Teknol.*, pp. 237–241, 2020.
- [13] J. Hao and T. K. Ho, "Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language," *J. Educ. Behav. Stat.*, vol. 44, no. 3, pp. 348–361, 2019, doi: 10.3102/1076998619832248.
- [14] D. K. Putra, I. Iwut, and R. D. Atmaja, "Simulasi Dan Analisis Speaker Recognition Menggunakan Metode Mel Frequency Cepstrum Coefficient (mfcc) Dan Gaussian Mixture Model (gmm)," *eProceedings Eng.*, vol. 4, no. 2, pp. 1766–1772, 2017, [Online]. Available: <http://libraryproceeding.telkomuniversity.ac.id/index.php/engineering/article/view/487/460>.
- [15] A. J. Bowers and X. Zhou, "Receiver Operating Characteristic (ROC) Area Under the Curve (AUC): A Diagnostic Measure for Evaluating the Accuracy of Predictors of Education Outcomes," *J. Educ. Students Placed Risk*, vol. 24, no. 1, pp. 20–46, 2019, doi: 10.1080/10824669.2018.1523734.