Terbit online pada laman web jurnal: **http://journal.ittelkom-pwt.ac.id/index.php/dinda**

JURNAL DINDA

Kelompok Keahlian Rekayasa Data
Institut Teknologi Telkom Purwokerto

Vol. 1 No. 1 (2021) ISSN Media Elektronik: 12345-XYZ

Pengenalan Jenis Kelamin Manusia Berbasis Suara Menggunakan MFCC dan GMM

Faisal Dharma Adhinata[1], Diovianto Putra Rakhmadani[2], Alon Jala Tirta Segara[3]
[1,2,3]Rekayasa Perangkat Lunak, Fakultas Informatika, Institut Teknologi Telkom Purwokerto
[1]faisal@ittelkom-pwt.ac.id, [2]diovianto@ittelkom-pwt.ac.id, [3]alon@ittelkom-pwt.ac.id

Abstract

Biometric information that exists in humans is unique from one human to another. One of the biometric data that is easily obtained is the human voice. The human voice is identic data that can differentiate between individuals. When we hear human voices directly, it is easy for our ears to tell the person who is speaking is male or female. But sometimes male voices can resemble girls and vice versa. Therefore, we propose a human voice detection system through Artificial Intelligence (AI) in machine learning. In this study, we used the Mel Frequency Cepstrum Coefficients (MFCC) method to extract human voice features and Gaussian Mixture Models (GMM) for the classification of female or male voice data. The experiment results showed that the system built was able to detect human gender through biometric voice data with an accuracy of 81.18%.

Keywords: voice, gender, machine learning, MFCC, GMM

Abstrak

Informasi biometric yang ada pada manusia sangat unik antara manusia yang satu dengan lainnya. Salah satu informasi biometric yang mudah diperoleh adalah suara manusia. Suara manusia merupakan data identic yang mampu membedakan antar individu. Apabila mendengar suara manusia secara langsung, telinga kita mudah membedakan yang berbicara itu berjenis kelamin laki-laki atau perempuan. Namun terkadang suara laki-laki bisa menyerupai perempuan, begitu juga sebaliknya. Oleh karena itu, kami mengusulkan sistem deteksi suara manusia melalui Artificial Intelligence (AI) pada bidang machine learning. Pada penelitian ini kami menggunakan metode Mel Frequency Cepstrum Coefficients (MFCC) sebagai ekstraksi fitur suara manusia dan Gaussian Mixture Models (GMM) untuk klasifikasi data suara perempuan atau laki-laki. Hasil percobaan menunjukkan sistem yang dibangun mampu mendeteksi jenis kelamin manusia melalui data biometric suara dengan akurasi 81,18%.

Kata kunci: suara, jenis kelamin, machine learning, MFCC, GMM

1. Pendahuluan

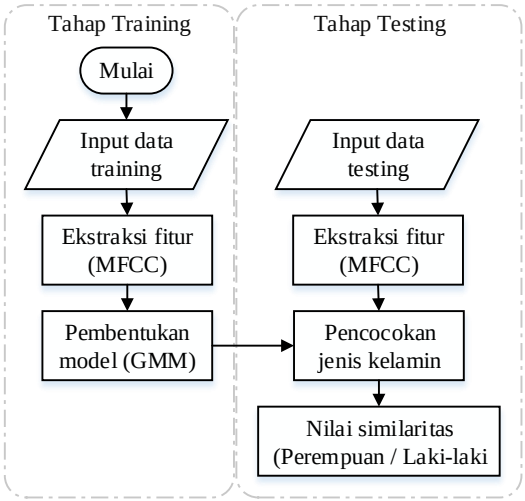
Metode konvensional yang masih sering digunakan untuk mengenali identitas seseorang biasanya menggunakan ID Card (KTP, SIM, paspor, dsb). Metode pengenalan konvensional memiliki keterbatasan, yaitu mudah rusak dan hilang [1]. Sistem pengenalan biometric mampu mengatasi keterbatasan ini karena sistem identifikasi menggunakan bagian tubuh manusia. Biometric adalah teknik yang mempelajari fisik atau tingkah laku manusia yang sering digunakan sebagai input pengenalan pola [2]. Keunggulan biometrik dalam hal otentifikasi adalah kemampuannya membedakan satu individu dengan individu lain secara akurat, sulit diduplikasi, dan tidak mudah hilang [3]. Informasi biometric yang paling mudah didapat dan sering digunakan dalam kehidupan sehari-hari adalah suara.

Pengenalan gender merupakan hal penting dalam proses keamanan. Misalnya dalam sebuah acara yang acaranya hanya boleh didatangi oleh kaum perempuan, sangat perlu dilakukan pengenalan gender sebagai keamanan. Sekarang banyak kasus kaum laki-laki menyamar sebagai perempuan dengan memakai cadar,

Diterima Redaksi : 11-12-2020 | Selesai Revisi : 21-12-2021 | Diterbitkan Online : 02-02-2021

sehingga tidak terlihat wajahnya. Bahkan sering orang laki-laki suaranya bisa dibuat menyerupai perempuan, begitu juga sebaliknya. Oleh karena itu, salah satu pengenalan gender dapat melalui suara seseorang yang setiap orang mempunyai karakter suara masing-masing. Sampel suara manusia mengandung banyak informasi. Pada penelitian ini melakukan analisis suara menggunakan machine learning yang dapat mendeteksi gender (jenis kelamin) orang melalui input suara orang. Metode pengenalan jenis kelamin melalui suara pada penelitian ini menggunakan MFCC sebagai ekstraksi fitur dan GMM sebagai pengklasifikasi.

2. Metode Penelitian



Gambar 1. Arsitektur Sistem Klasifikasi Jenis Kelamin Manusia

Penelitian ini menggunakan metode Mel Frequency Cepstrum Coefficients (MFCC) sebagai ekstraksi fitur suara dan Gaussian Mixture Models (GMM) untuk klasifikasi jenis kelamin perempuan atau laki-laki. Gambar 1 menunjukkan arsitektur sistem yang diusulkan. Pada tahapan MFCC terdapat proses preprocessing sinyal suara, yaitu tahap pre-emphasis. Pre-emphasis berfungsi untuk menstabilkan nilai magnitude dari sinyal suara. Sinyal suara pada data training akan dilakukan ekstraksi fitur menggunakan MFCC, masing-masing sinyal suara (suara perempuan dan laki-laki) mempunyai informasi fitur sendiri-sendiri. Kemudian hasil ekstraksi fitur dimodelkan dengan GMM. Selanjutnya dilakukan testing menggunakan suara orang yang berbeda untuk mengetahui jenis kelamin orang yang mengucapkan suara tersebut. Data testing suara dilakukan ekstraksi fitur terlebih dahulu menggunakan MFCC. Kemudian model data training hasil dari GMM digunakan untuk menghitung nilai fitur pada kedua model (perempuan dan laki-laki). Model yang menghasilkan nilai maksimum adalah prediksi jenis kelamin dari data testing.

2.1. Dataset

Dataset penelitian ini menggunakan data training dan data testing. Data yang digunakan dalam percobaan ada semuanya suara orang dewasa.

a. Data training

Data training menggunakan rekaman suara berformat .wav yang terdiri 5 suara laki-laki dan 5 suara perempuan.

b. Data testing

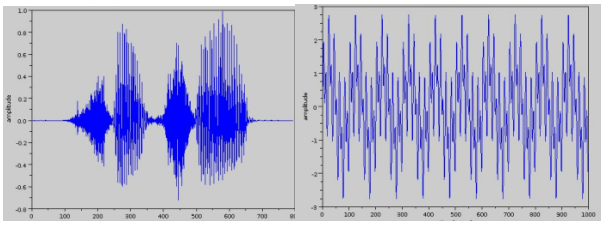
Data testing menggunakan rekaman suara rekaman dari AudioSet yang diambil dari https://research.google.com/audioset.

2.2. Frame suara

Sinyal suara berupa urutan angka yang menunjukkan amplitude dari suara yang diucapkan oleh manusia. Terdapat 3 konsep dalam pengolahan sinyal suara :

a. Framing

Suara manusia adalah sinyal non-stationary, yang fekruensinya berubah – ubah terhadap waktu. Untuk kebutuhan analisis sinyal suara, perlu dilakukan perubahan ke sinyal stationary. Proses perubahan dapat dilakukan dengan membagi sinyal suara menjadi frame pendek 20 sampai 30 ms. Framing adalah membagi sinyal suara menjadi beberapa frame yang bertujuan untuk memudahkan dalam perhitungan dan melakukan analisis sinyal, satu frame terdiri dari beberapa sampel tergantung pada tiap berapa detik suara akan disampel dan berapa besar frekuensi samplingnya [4]. Gambar 2 menunjukkan perbedaan sinya suara non-stationary dan stationary.

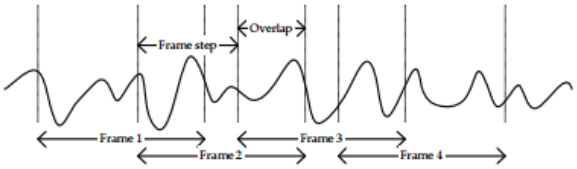


a b

Gambar 2. Sinyal suara, a) sinyal non-stationary b) Sinyal stationary

b. Overlapping

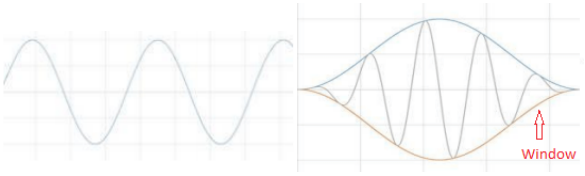
Overlapping adalah proses yang diawali dengan membandingkan panjang data yang akan dioverlapping dari masing-masing frame dari beberapa frame yang akan digabungkan. Setelah itu akan dibandingkan dan dicari mana yang paling pendek dari keduanya akan dijadikan patokan dalam penentuan banyak data yang akan diproses overlapping. Kehilangan sampel pada awal dan akhir frame menyebabkan kesalahan pada representasi frekuensi. Untuk itu perlu dilakukan overlapping frame yang pada umumnya antara 10-15 ms. Gambar 3 menunjukkan overlapping frame pada sinyal suara.



Gambar 3. Overlapping frame pada sinyal suara

c. Windowing

Dalam melakukan pemrosesan sinyal, input yang dimasukkan akan terbentuk sinyal yang amplitudonya bervariasi pada awal maupun akhir frame. Hal tersebut dapat menghambat pemrosesan sinyal dan menghasilkan keluaran yang kurang akurat. Untuk itu perlu diaplikasikan suatu window penghalus pada setiap melakukan overlapping antara satu frame dengan frame yang lain, sehingga dapat dibangkitkan suatu fitur yang lebih halus sepanjang durasi tersebut. Fungsi window dilakukan dengan mengecilkan amplitudo sampai angka 0 pada akhir sinyal.



a b

Gambar 4. a) sinyal stationary b) windowing sinyal

2.3. Ekstraksi Fitur menggunakan MFCC

Setelah dilakukan ekstraksi pada frame suara, tahap selanjutnya adalah mendapatkan fitur menggunakan MFCC untuk setiap frame suara. MFCC mengubah nilai sinya suara kedalam domain cepstral. Secara teori, ucapan diasumsikan sebagai sumber konvolusi (udara yang dikeluarkan dari paru-paru) dan filter (saluran suara manusia). Tujuan dari metode ini adalah untuk melakukan ektraksi pada filter dan menghapus bagian sumber. Tahapan MFCC ditunjukkan pada Gambar 5.

a. Pre-emphasis

Pre-emphasis berfungsi untuk menstabilkan nilai magnitude dari sinyal suara. Persamaan pre-emphasis ditunjukkan pada persamaan (1).

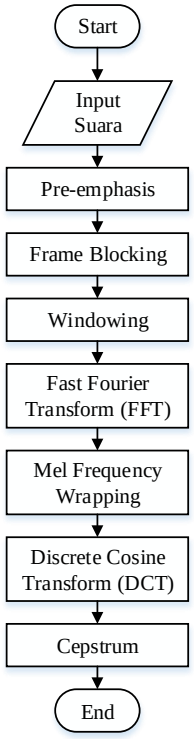
$$s'_n = s_n - \alpha s_{n-1} \quad (1)$$

dimana:

s_n : nilai sampel ke-n
α : konstanta pre-emphasis, $0.9 \leq \alpha \leq 1.0$

b. Frame Blocking

Framing berfungsi membagi sinyal suara menjadi beberapa frame dengan panjang sampel tertentu. Sinyal suara yang terdiri dari S sampel $(X(S))$ akan dibagi menjadi beberapa frame yang berisi N sampel, masing-masing frame akan dipisahkan oleh $M(M < N)$ [5].



Gambar 5. Tahapan metode MFCC

c. Windowing

Windowing berfungsi meredam noise yang muncul di kedua ujung frame. Proses windowing yang dipakai pada penelitian ini adalah proses Hamming Window, proses tersebut dapat dituliskan dalam persamaan (2) [6].

$$w(n) = 0.54 - 0.46\, cos = \frac{2\pi n}{N-1} \quad (2)$$

Dimana:

N : Jumlah sample pada masing-masing frame
n : $0,1,2,3,\dots,N-1$

d. FFT

Fast Fourier Transform (FFT), diguanakan untuk mengubah domain waktu ke domain frekuensi. Proses FFT dapat dituliskan dalam persamaan (3) [7].

$$f(n) = \sum_{K=0}^{N-1}(Y_k)e^{-\frac{2\pi jkn}{N}}, \quad n = 0,1,2,\dots,N-1 \quad (3)$$

Dimana:

$f(n)$: Frekuensi
k : $0,1,2,\dots,(N-1)$
N : Jumlah sample pada masing-masing frame
j : bilangan imajiner
n : $1,2,3,\dots,(N-1)$

e. Mel-frequency wrapping

Pada tahap ini, sinyal suara pada domain frekuensi diubah menjadi domain frekuensi mel. Persamaan

untuk menghitung skala mel pada frekuensi dalam Hz sebagaimana persamaan (4).

$$Mel\, f = 2595 * log_{10}(1 + \frac{f}{700}) \quad (4)$$
$$f = 700(10^{\frac{m}{2595}} - 1)$$

dimana:
$Mel\, f$: nilai frekuensi mel dari f
Tahap ini menghasilkan sejumlah mel filter bank. Nilai mel filter bank merepresentasikan seberapa besar energi pada rentang frekuensi yang ada pada masing-masing filter mel [3].

f. DCT
DCT (Discrete Cosine Transform) untuk mengubah ke dalam domain waktu. Persamaan DCT ditunjukkan dalam persamaan (5) [8].

$$C_n = \sum_{k=1}^{K}(logS_k)\cos\left[n\left(k-\frac{1}{2}\right)\frac{\pi}{k}\right], \quad n = 1,2,\dots,k \quad (5)$$

Dimana:
C_n : koefesien cepstrum mel-frequency
S_k : mel frekuensi
n : bilangan bulat dari $1,\dots,N$ (jumlah total sampel)
k : jumlah coefficient

Proses ini akan menghasilkan log dari perkalian DCT yang sudah diubah ke domain waktu. Hasil log perkalian domain waktu ini menghasilkan mel-frequency cepstrum coefficient (MFCC) [9].

2.4. Training Gender menggunakan GMM
Untuk melakukan pengenalan terhadap jenis kelamin (gender) dari data ekstraksi fitur, maka dimodelkan kedua gender (perempuan dan laki-laki) menggunakan metode GMM. GMM adalah model klastering probabilitas untuk merepresentasikan ada tidaknya sub populasi dari semua populasi [10]. GMM berfungsi untuk memodelkan sejumlah data menjadi sebuah distribusi Gaussian dengan parameter mean μ dan variance σ^2 tertentu. Mean μ adalah titik pusat dari distribusi Gaussian sedangkan variance σ^2 adalah ukuran persebaran nilai pada set data. Kelebihan metode GMM adalah mampu memodelkan lebih dari satu Gaussian untuk sebuah set data. Konsep training pada GMM menggunakan distribusi probabilitas dari sebuah kelas dengan kombinasi linear k distribusi Gaussian atau yang dikenal sebagai komponen GMM. Nilai probabilitas pada data (fitur vector) didefinisikan dengan persamaan (6).

$$P(X|\lambda) = \sum_{k=1}^{K} w_k P_k(X|\mu_k, \sum_k) \quad (6)$$

dimana $P_k(X|\mu_k, \sum_k)$ adalah Gaussian distribution

$$P_k(X|\mu_k, \sum_k) = \frac{1}{\sqrt{2\pi|\sum_k|}} e^{\frac{1}{2}(X-\mu_k)^T\sum^{-1}(X-\mu_k)}$$

Data training X_i dari kelas λ digunakan untuk menentukan parameter rata-rata μ, matriks co-variance $\sum$ dan bobot w pada komponen k. Tahap pertama adalah mengidentifikasi kluster k pada data menggunakan algoritma K-means dan memberikan bobot yang sama $w = \frac{1}{k}$ pada masing-masing kluster. Gaussian distribution k dipasangkan dengan kluster k. Parameter μ, σ, dan w dari semua kluster diupdate pada setiap iterasi sampai konvergen. Metode yang sering digunakan dalam hal ini adalah algoritma Expectation Maximization (EM). Algoritma EM adalah sebuah prosedur iteratif untuk menghitung estimasi Maximum Likelihood (ML) yang muncul pada hidden data. Pada estimasi ML, kita ingin menghitung model parameter dengan kemungkinan paling besar bagi data yang terobservasi.

2.5. Evaluasi menggunakan data testing
Data testing menggunakan rekaman dari AudioSet yang diambil dari https://research.google.com/audioset. Data testing dilakukan ekstraksi fitur dengan ukuran frame 25 ms dan overlap antar frame 10 ms. Selanjutnya nilai log-likelihood pada masing-masing frame setiap sampel, $x_1, x_2, \dots, x_i$ pada masing-masing jenis kelamin yaitu $P(x_i|perempuan)$ dan $P(x_i|laki-laki)$ dilakukan perhitungan. Menggunakan persamaan Gaussian distribution, log-likelihood pada frame suara wanita dilakukan perhitungan dengan mengganti μ dan $\sum$ pada model GMM perempuan. Hal yang sama juga diterapkan pada model GMM laki-laki.

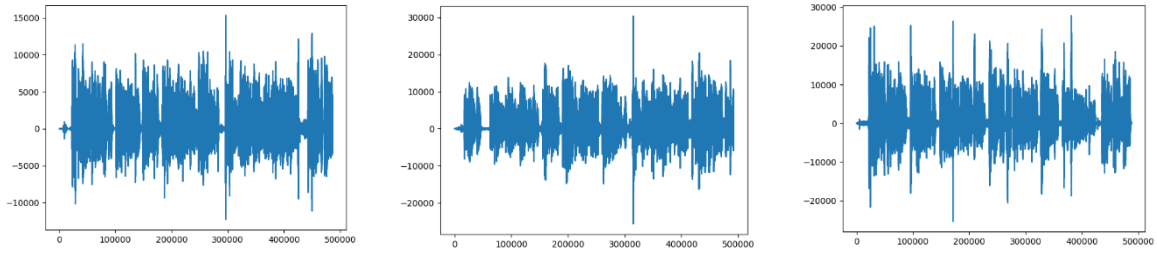
3. Hasil dan Pembahasan

Percobaan ini menggunakan sampel suara mahasiswa Magister Ilmu Komputer UGM sebagai data training. Sampel yang diambil berjumlah 10 suara, 5 suara perempuan dan 5 suara laki-laki. Rekaman suara dilakukan dengan membaca berita pada internet dan direkam menggunakan smartphone dengan sample rate 16 kHz dan disimpan menggunakan format (.wav). Perekaman suara untuk data training masing-masing berdurasi 30 detik.

Selanjutnya untuk data testing menggunakan rekaman dari AudioSet yang diambil dari https://research.google.com/audioset dengan jumlah data suara perempuan 550 data dan suara laki-laki 550 data. Semua data berdurasi 10 detik dan menggunakan sample rate 16 kHz. Rekaman data testing menggunakan Bahasa Inggris untuk menguji tingkat akurasi metode MFCC dan GMM ketika dimatchkan dengan data training Bahasa Indonesia.

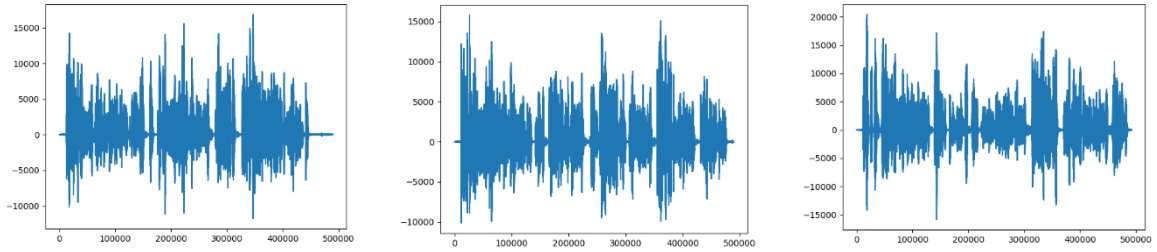
3.1. Sinyal suara

Suara manusia antara individu yang satu dengan yang lainnya memiliki perbedaan. Namun demikian dalam perbedaan-perbedaan tersebut masih terdapat adanya kesamaan antara suara perempuan yang satu dengan perempuan yang lain dan antara laki-laki yang satu dengan laki-laki yang lain. Gambar 6 dan 7 merupakan hasil gelombang sinyal suara perempuan dan laki-laki.



Data training 1 Data training 2 Data training 3

Gambar 6. Gelombang sinyal suara data training perempuan



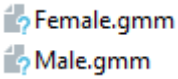
Data training 1 Data training 2 Data training 3

Gambar 7. Gelombang sinyal suara data training laki-laki

Karakteristik suara perempuan dan laki-laki apabila didengarkan dengan seksama, suara perempuan cenderung lebih tinggi dibanding suara laki-laki. Melalui perbedaan tersebut, bisa dilakukan identifikasi terhadap suara manusia dengan analisis perbandingan frekuensi suaranya untuk klasifikasi suara perempuan atau laki-laki.

3.2. Pemodelan data training

Pada tahap training, hasil ekstraksi fitur metode MFCC dilakukan proses training menggunakan GMM yang masing-masing menghasilkan model perempuan dan laki-laki dengan file ekstensi (.gmm). File hasil pembuatan model ditunjukkan pada Gambar 8.



Gambar 8. Hasil pemodelan sinyal suara perempuan dan laki-laki

GMM merepresentasikan setiap distribusi dari jumlah bobot komponen M (mixture) Gaussian. Pada percobaan ini menggunakan variasi mixture GMM, yaitu 4, 8, dan 16. Pada tahap testing, hasil ekstraksi fitur data testing dibandingkan dengan model gender pada tahap training. Sebagai hasil pengenalan suara perempuan atau laki-laki, kriteria log-likelihood menunjukkan bagaimana data testing di*match*kan terhadap model hasil training. Nilai terbesar pada log-likelihood sebagai hasil pengenalan gender pada percobaan ini.

3.3. Matching feature data testing

Hasil percobaan menggunakan 1100 data suara manusia yang terdiri dari 550 suara laki-laki dan 550 suara perempuan disajikan pada Tabel 1.

Tabel Hasil pecobaan dengan variasi mixture GMM

Mixture GMM	Hasil Klasifikasi Gender		Akurasi
	Benar	Salah	
4	828	272	75,27%
8	865	235	78,63%
16	893	207	81,18%

Berdasarkan Tabel diatas terlihat semakin banyak mixture pada GMM akurasi klasifikasi gender semakin meningkat. Akurasi terbaik diperoleh dengan mixture GMM 16 yaitu 81.18%.

4. Kesimpulan

Penelitian ini memberi gambaran klasifikasi jenis kelamin melalui data biometric manusia yaitu suara. Salah satu manfaat pengenalan suara adalah sebagai keamanan dari tindak kejahatan, terutama bagi kaum perempuan. Penggunaan kombinasi metode MFCC

sebagai ekstraksi fitur dan GMM sebagai klasifikasi data suara laki-laki dan perempuan, sistem yang dibangun menghasilkan akurasi 81,18%. Penelitian kedepan dapat memodifikasi atau menggunakan kombinasi metode lain untuk memperoleh hasil akurasi yang lebih akurat.

Daftar Rujukan

[1] A. A. Andarinny, C. E. Widodo, and K. Adi, "Perancangan Sistem Identifikasi Biometrik Jari Tangan Menggunakan Laplacian Of Gaussian dan Ektraksi Kontur," *Youngster Phys. J.*, vol. 6, no. 4, pp. 304–314, 2017.

[2] F. T. Elektro, U. Telkom, and D. Tree, "Deteksi Kepribadian Anak Dengan Sidik Jari Menggunakan Metode K- Nearest Neighbor (Knn) Dan Decision Tree Detection of Children ' S Personality With Fingerprint Using K-Nearest Neighbor (Knn) and Decision Tree Methods," vol. 6, no. 2, pp. 5549–5556, 2019.

[3] 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://libraryeproceeding.telkomuniversity.ac.id/index.php/engineering/article/view/487/460.

[4] D. T. Handoko and P. Kasih, "Voice Recognition untuk Sistem Keamanan PC Menggunakan Metode MFCC dan DTW," *Gener. J.*, vol. 2, no. 1, pp. 57–68, 2018, doi: 10.29407/gj.v2i1.12058.

[5] M. Azizah, A. Hidayatno, and Y. Christyono, "APLIKASI PENGENAL PENGUCAP BERBASIS IDENTIFIKASI SUARA DENGAN EKSTRAKSI CIRI MEL-FREQUENCY CEPSTRUM COEFFICIENTS (MFCC) DAN KUANTISASI VEKTOR Mega Tiara Nur Azizah *) , Achmad Hidayatno , and Yuli Christyono Abstrak Pendahuluan Metode," *Transient*, vol. 6, no. 4, pp. 639–643, 2017.

[6] A. K. Munggaran, E. C. Djamal, and R. Yuniarti, "Identifikasi Suara Pengontrol Lampu Menggunakan Mel-Frequency Cepstral Coefficients dan Hidden Markov Model," *Semin. Nas. Apl. Teknol. Inf. 2017*, pp. 17–22, 2017.

[7] A. Syaifuddin and S. Suryono, "Fast Fourier Transform (Fft) Untuk Analisis Sinyal Suara Doppler Ultrasonik," *Youngster Phys. J.*, vol. 3, no. 3, pp. 181–188, 2014.

[8] I. P. Prawiro, R. Magdalena, and I. N. A. Ramatryana, "SIMULASI DAN ANALISIS PERBANDINGAN ANTARA METODE DISCRETE COSINE TRANSFORM (DCT) DAN MODIFIED DISCRETE COSINE TRANSFORM (MDCT) PADA PEMISAHAN REFF LAGU," vol. 5, no. 3, pp. 5513–5520, 2018.

[9] S. Helmiyah, A. Fadlil, and A. Yudhana, "Pengenalan Pola Emosi Manusia Berdasarkan Ucapan Menggunakan Ekstraksi Fitur Mel-Frequency Cepstral Coefficients (MFCC)," *CogITo Smart J.*, vol. 4, no. 2, p. 372, 2019, doi: 10.31154/cogito.v4i2.129.372-381.

[10] J. W. G. Putra, "Pengenalan Konsep Pembelajaran Mesin dan Deep Learning," vol. 1.4, pp. 1–235, 2020, [Online]. Available: https://www.researchgate.net/publication/323700644.