Simulasi Ekstraksi Fitur Suara menggunakan Mel-Frequency Cepstrum Coefficient

Ratna Hartayu[1)], Santoso[2)], Choirul Anam[3)], Dimas Abdul Aziz[4)]
[1)2)3)4)]Program Studi Teknik Elektro, Fakultas Teknik, Universitas 17 Agustus 1945 Surabaya
Jl. Semolowaru No.45, Menur Pumpungan, Kec. Sukolilo, Kota SBY, Jawa Timur 60118
[1)]rhartayu@untag-sby.ac.id
[2)]santoso@untag-sby.ac.id
[3)]Choirul@gmail.com
[4)]Aziz@gmail.com

Abstrak

Dalam penelitian ini metode *Mel Frequency Cepstrum Coefficients* (MFCC) digunakan untuk menghasilkan ekstraksi ciri dari suara. Untuk mengimplementasikan sistem ini digunakan rekaman suara berupa kata perintah. Menggunakan 9 kata perintah, dengan 3 intonasi berbeda dan melibatkan 100 orang dalam menghasilkan data suara. Dalam prosesnya sinyal suara diolah menggunakan filter hamming window, *fast fourier transform* (FFT), IFFT. Penelitian ini menghasilkan analisa simulasi perubahan data sinyal suara, *spectrum, spectogram, logspectogram,* dan *MFCC,* menggambarkan implementasi pengenalan pola suara, sehingga merupakan sumber masukan *dataset* untuk penelitian klasifikasi suara berikutnya.

Kata kunci — Simulasi, Windowing, Ektraksi, MFCC

Abstract

In this study, the Mel Frequency Cepstrum Coefficients (MFCC) method was used to produce feature extraction from sound. To implement this system, a voice recording is used in the form of command words. Using 9 command words, with 3 different intonations and involving 100 people in generating voice data. In the process, the voice signal is processed using a hamming window filter, fast fourier transform (FFT), IFFT. This research produces a simulation analysis of changes in voice signal data, spectrum, spectogram, logspectogram, and MFCC, describing the implementation of speech pattern recognition, so that it is a dataset input source for the next sound classification research.

Keywords *- Simulation, Windowing, Extraction*

1. PENDAHULUAN

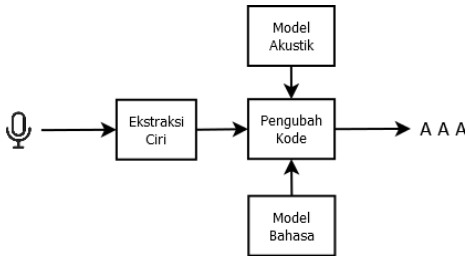
Perkembangan teknologi pengenalan suara telah membuat aktifitas kehidupan manusia semakin meningkat. Saat ini perkembangannya mampu meningkatkan kualitas maupun kuantitas kegiatan diberbagai industri. Teknologi pengolahan suara juga telah menjangkau sisi hiburan dan pendidikan bagi manusia. Salah satu cara menganalisa suara dengan menambah metode pengenalan suara (speech recognition). Speech recognition dapat diperoleh dengan metode ekstraksi ciri yang dapat digunakan untuk mengetahui identitas sinyal audio, yaitu; Mel-Frequency Cepstrum Coefficient (MFCC)[1]–[3].

DOI: 10.34128/jsi.v8i1.357 *Received: 24 Oktober 2021*
Accepted: 4 Juli 2022

2. TINJAUAN PUSTAKA

2.1. Speech recognition

Speech recognition suatu teknik yang menerima masukan berupa kata yang diucapkan. Teknologi ini mampu mengenali dan memahami kata-kata dengan cara digitalisasi kata dan mencocokkan sinyal digital tersebut dengan suatu pola tertentu yang tersimpan dalam suatu perangkat. Kata-kata yang diucapkan diubah bentuknya menjadi sinyal digital dengan cara mengubah gelombang suara menjadi sekumpulan angka kemudian disesuaikan dengan kode-kode tertentu untuk mengidentifikasikan kata-kata tersebut. Hasil dari identifikasi kata dapat ditampilkan dalam bentuk tulisan atau dapat dibaca oleh perangkat teknologi sebagai komando untuk melakukan suatu pekerjaan[4], [5].

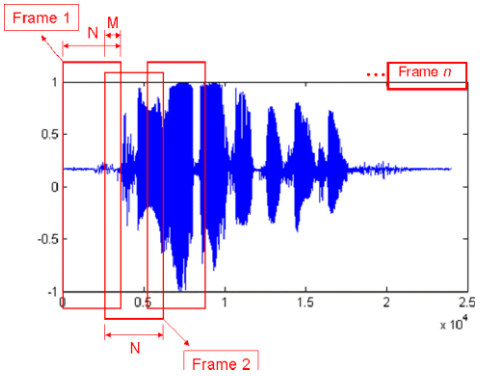


Gambar 1 Blok diagram Speech Recognition.

Gambar 1 menunjukkan model Speech Recognition dimana suara melalui mikrophone akan mengubah suara menjadi sinyal digital, kemudian diekstraksi cirinya, diproses untuk menjadi sebuah teks.

2.2. Frame blocking

Sinyal suara X yang dari S sample (X(S)), dibagi menjadi beberapa frame yang berisi N sample, masing-masing frame dipisahkan oleh M (M<N). Frame pertama berisi sampel N pertama. Frame kedua dimulai M sampel setelah permulaan frame pertama, sehingga frame kedua ini overlap terhadap fram pertama sebanyak N-M sampel. Seterusnya, frame ketiga dimulai M sampel setelah frame kedua (juga overlap sebanyak N-M sampel terhadap frame kedua). Proses ini berlanjut sampai seluruh suara tercakup dalam frame. Hasil dari proses ini adalah matriks dengan N baris dan beberapa kolom sinyal X[N][6].



Gambar 2. Frame blocking terhadap sinyal suara

Pada gambar 2, Sinyal *S* disekat kedalam urutan frame *n* yang tumpah tindih, F_1, F_2,..., dan F_n di representasikan kedalam

$$S = (F_1, F_2, \dots F_n) \quad (1)$$

Diasumsikan bahwa lama waktu *frame* dan tumpang tindih *frame* untuk *S* adalah *N* dan *M*, bingkai F_1 mengandung Sampel *N*, yang mana ditandai sebagai

$$F_1 = (x_1, x_2, \dots x_n) \tag{2}$$

Dimana x_1 adalah sampel pertama dan x_n adalah sampel ke *N* dalam F_1, oleh karena itu frame ke *n* adalah ditentukan oleh

$$F_n = \left(x_{(n-1)(N-M)+1}, x_{(n-1)(N-M)+2}, \dots x_{(n-1)(N-M)+N}\right) \tag{3}$$

Oleh karena itu persamaan (2) dapat ditulis

$$S = \left(x_1, \dots, x_i, \dots, x_{(n-1)(N-M)+N}\right) \tag{4}$$

Dimana x_i menunjukkan sebuah sampel ucapan, $i = 1,2, \dots, (n-1)(N-M)+N$

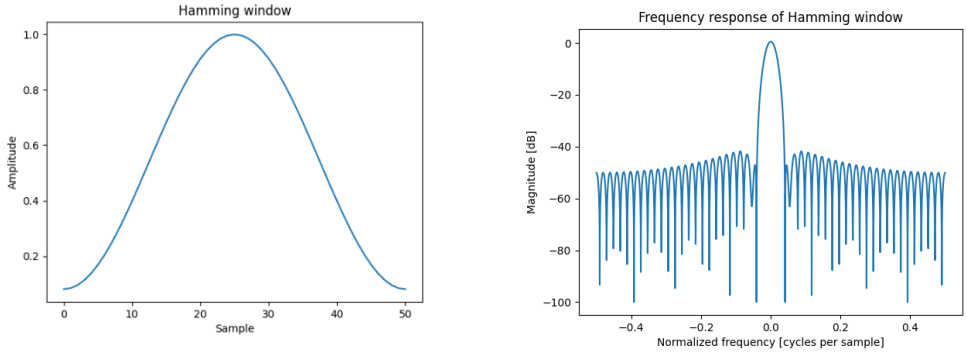
Dalam phase implementasi, durasi frame dan tumpangtindih frame dipasang, secara umum masing-masing 20-30ms(mili second) dan separuh durasi frame.

2.3. Windowing

Sinyal analog yang telah diubah menjadi sinyal digital dibaca *frame* demi *frame* dan pada setiap frame dilakukan *windowing* dengan fungsi tertentu. Proses *windowing* bertujuan untuk meredam *noise* di awal dan akhir setiap *frame*. Dimana rumus umum yang digunakan adalah *Hamming Window*,

$$H(n) = 0.54 - 0.46\cos(2\pi\frac{n}{N-1}),\ 0 \le n \le N-1 \tag{5}$$

dimana *N* adalah banyaknya sample pada tiap frame dan *n* adalah bilangan bulat dari *0* hingga *N-1*. [5].



Gambar 3. (a) Jendela hamming, (b) Log magnitude Respon frekuensi

Gambar 3a dan 3b menunjukkan respon waktu dan respon spektral filter. Lebar lobus utama cocok dengan hann, bahwa sidelobe tertinggi dilemahkan terhadap lobus utama sebesar 43 dB, dan tingkat atenuasi asimtotik adalah 6dB/oktaf[8].

2.4. Fast Fourier Transform (FFT)

FFT adalah algoritma cepat dari Discrete Fourier Transform (DFT) yang berguna untuk mengubah setiap frame dengan sampel N dari domain waktu menjadi domain frekuensi, sebagaimana didefinisikan sebagai berikut:

$$X_n = \sum_{k=0}^{N-1} \binom{n}{k} X_k e^{-2\partial jkn/N} \quad n=0,1,2 \dots n^{-1}\ dan\ n=\sqrt{-1} \tag{7}$$

Hasil dari tahapan ini biasanya disebut sebagai spektrum atau periodogram[9], [10].

2.5. Mel-frequency

Skala frekuensi mel adalah frekuensi rendah yang linier di bawah 1000 Hz dan frekuensi tinggi yang logaritmik di atas 1000 Hz. Persamaan berikut menunjukkan hubungan skala mel dengan frekuensi dalam Hz:

$$f = \begin{cases} 2595 * log10\left(1 + \frac{F_{Hz}}{700}\right) & F_{Hz} > 1000 \\ F_{Hz} & F_{Hz} < 100 \end{cases} \quad (8)$$

$$M(f) = 1125\ x \ln\left(1 + \frac{f}{700}\right) \quad (9)$$

$$Mel_i = Mel_{bawah} + \frac{i(Mel_{atas} - Mel_{bawah})}{n+1} \quad (10)$$

$$M^{-1}(m) = 700\ x(e^{\frac{m}{1125}} - 1) \quad (11)$$

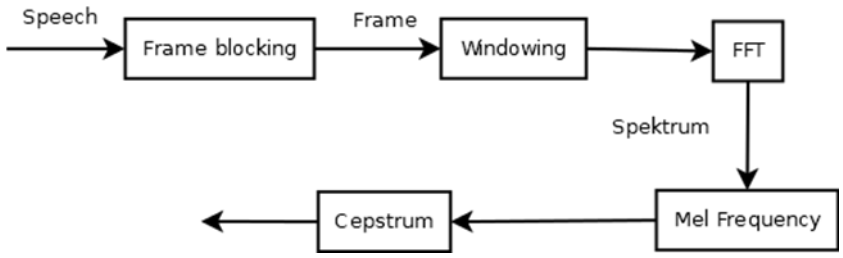
Setelah diperoleh sejumlah titik, titik tersebut diubah kembali kedalam satuan frekuensi dengan menggunakan persamaan (8). Untuk membuat filter MFCC digunakan persamaan (9), (10), (11) untuk mendapatkan nilai *filterbank*.

$$H_m(k) = \begin{cases} 0, & k < f(m-1) \quad (12) \\ \dfrac{k - f(m-1)}{f(m) - f(m-1)}, & f(m-1) \le k \le f(m) \quad (13) \\ \dfrac{f(m+1) - k}{f(m+1) - f(m)}, & f(m) \le k \le f(m+1) \quad (14) \\ 0, & k > f(m+1) \quad (15) \end{cases}$$

Transformasi data suara dikalikan dengan *filterbank*, hasil perkalian ditransformasi dengan *Discrete Cosine Transform*(DCT) dan dinormalisasi. DCT berfungsi untuk mendekorelasi koefisien hasil dari *filterbank* skala mel[11], [12]. Filterbank didesain tumpah tindih, Setiap segitiga yang terbentuk adalah *window* yang diterapkan pada representasi frekuensi suara. Penerapan setiap window ke energi FFT akan menghasilkan spektrum Mel.

3. METODE PENELITIAN

Penelitian ini menggunakan 9 kata perintah seperti “Aktif”, “Non aktif”, “Berhenti”, “Hidup”, “Kanan”, “Kiri”, “Maju”, “Mundur”, “Robot”, perekaman suara dilakukan menggunakan program perekam suara Moo0, dengan durasi perekaman 2-3 detik. Gambar 4 menunjukkan proses ekstraksi, sinyal suara hasil rekaman diambil per frame untuk menghasilkan informasi sinyal audio sebesar N frame, kemudian menggunkana filter window hamming untuk menghapus gangguan.

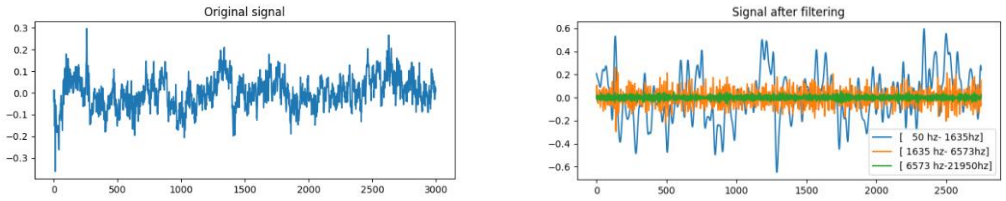


Gambar 4 Blok diagram MFCC

Setelah proses menghilangkan ganguan, dilakukan proses perubahan domain waktu terhadap sinyal kedalam domain frekuensi (*fast fourier transform*)yang berguna untuk memudahkan dalam proses analisis sinyal. Dengan proses Mel Frequency melalui filter bank didapat nilai energi dari frequency band tertentu dalam sinyal suara.

4. PEMBAHASAN

Hasil perekaman suara tersimpan dalam bentuk file wav, bentuk sinyal dari kata “Aktif”, terlihat seperti pada gambar 5.

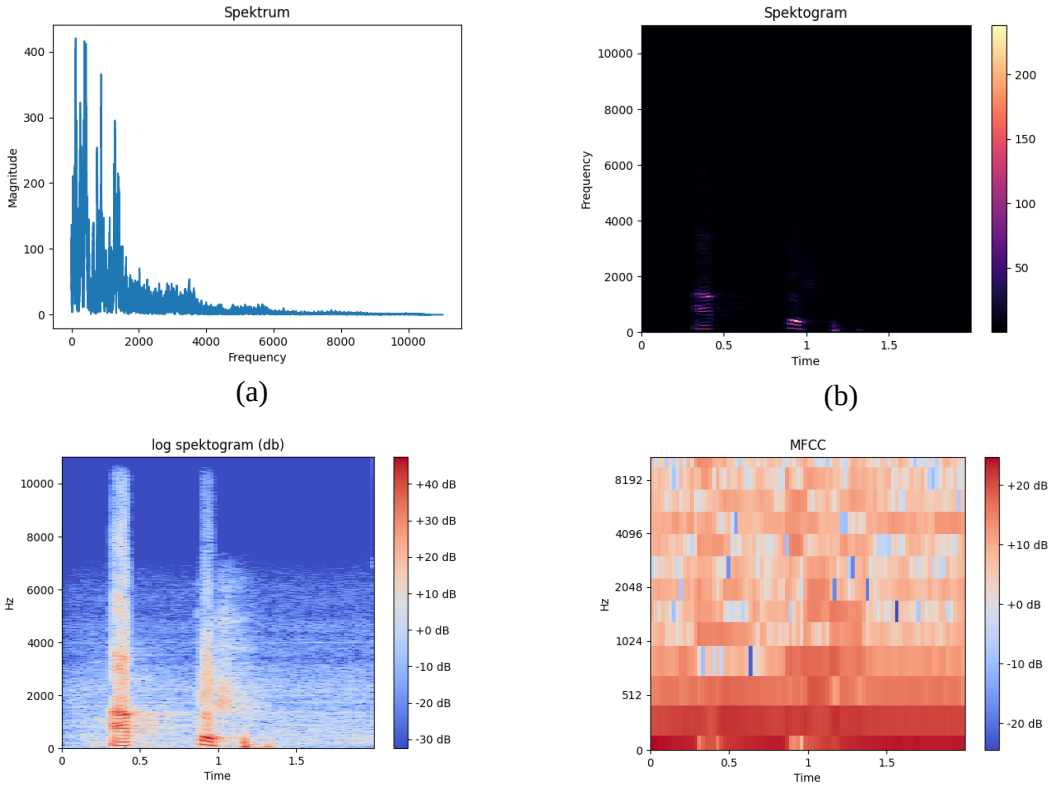


(a) (b)

Gambar 5a merupakan bentuk sinyal suara, gambar 5b bentuk sinyal sebelum dan sesudah filter dengan filter window hamming.

Pengujian menggunakan variasi dengan batas frekuensi yang berbeda-beda yaitu 50-1635Hz, 1635-6573Hz, dan 6573-21950Hz. Pada gambar 5b Sample rate menunjukan nilai sinyal audio yang diambil dalam tiga detik, ketika melakukan rekaman suara. Semakin tinggi nilai dari sample rate kualitas audio yang direkam akan semakin baik. Berdasarkan grafik pada Gambar 5b dapat disimpulkan bahwa akurasi terbaik dicapai sistem dengan menggunakan sample rate 50-1635Hz.
Berikutnya dilakukan ekstraksi fitur, menghasilkan spektrum, spectogram, log spektogram, dan MFCC, dimana sinyal suara, diproses menggunakan program python, menghasilkan bentuk seperti pada gambar 6.

p-ISSN: 2460-173X
e-ISSN: 2598-5841



Gambar 6. (a) Spektrum, (b) Spektrogram, (c) log spektogram, (d) MFCC

Gambar 6(a) merupakan nilai spektrum dari rekaman suara “aktif”, gambar 6(b) memperlihatkan pengubahan spektrum kedalam bentuk spektogram, bertujuan untuk melihat perubahan frekuensi terhadap waktu, yang divisualisasi dengan perubahan warna. Pada gambar 6(c) dibuat penyesuaian terhadap sumbu y (frekuensi) menjadi skala log, dan sumbu "warna" (amplitudo) menjadi Desibel, yang merupakan skala log amplitudo. Pada gambar 6(d) dalam skala MFCC dilakukan beberapa perhitungan transformasi non-linear dari skala frekuensi, dengan tujuan memberi informasi nilai db disepanjang frekuensi pada waktu yang sama

5. KESIMPULAN

Dalam penelitian ini kami telah berhasil mensimulasikan program guna menghilangkan gangguan noise dengan cut-off frekuensi 50-1635Hz pada sinyal suara dan mengekstraksi fitur suara dengan mendapatkan koefisien MFCC juga mempertimbangkan fungsi energi Delta sehingga dapat meningkatkan koefisien MFCC sesuai dengan kebutuhan. Terdapat penambahan kecepatan dan percepatan untuk mengekstrak koefisien MFCC. Teknik ekstraksi fitur MFCC lebih efektif. Fitur diekstraksi berdasarkan informasi yang disertakan dalam sinyal suara. Fitur yang diekstraksi disimpan dalam file .png. Penelitian berikutnya yaitu bagaimana hasil koefisien MFCC digunakan untuk dataset jaringan kecerdasan buatan, sehingga dapat dimanfaatkan untuk klasifikasi objek maupun kontrol instrumentasi.

UCAPAN TERIMA KASIH

Terimakasih kepada Universitas 17 Agustus 1945 Surabaya, telah membiayai penelitian ini

DAFTAR PUSTAKA

[1] S. Gupta, J. Jaafar, W. F. wan Ahmad, and A. Bansal, 'Feature Extraction Using Mfcc', *Signal Image Process. Int. J.*, vol. 4, no. 4, pp. 101–108, Aug. 2013, doi: 10.5121/sipij.2013.4408.

[2] D. Cao, X. Gao, and L. Gao, 'An Improved Endpoint Detection Algorithm Based on MFCC Cosine Value', *Wirel. Pers. Commun.*, vol. 95, no. 3, pp. 2073–2090, Aug. 2017, doi: 10.1007/s11277-017-3958-0.

[3] A. Setiawan, A. Hidayatno, and R. R. Isnanto, 'Aplikasi Pengenalan Ucapan dengan Ekstraksi Mel-Frequency Cepstrum Coefficients (MFCC) Melalui Jaringan Syaraf Tiruan (JST) Learning Vector Quantization (LVQ) untuk Mengoperasikan Kursor Komputer', *Transm. J. Ilm. Tek. Elektro*, vol. 13, no. 3, Art. no. 3, 2011, doi: 10.12777/transmisi.13.3.82-86.

[4] S. A. Alim and N. K. A. Rashid, *Some Commonly Used Speech Feature Extraction Algorithms*. IntechOpen, 2018. doi: 10.5772/intechopen.80419.

[5] J. Vanus *et al.*, 'Assessment of the Quality of Speech Signal Processing Within Voice Control of Operational-Technical Functions in the Smart Home by Means of the PESQ Algorithm', *IFAC-Pap.*, vol. 51, no. 6, pp. 202–207, Jan. 2018, doi: 10.1016/j.ifacol.2018.07.154.

[6] L.-M. Lee and F.-R. Jean, 'Adaptation of Hidden Markov Models for Recognizing Speech of Reduced Frame Rate', *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 2114–2121, Dec. 2013, doi: 10.1109/TCYB.2013.2240450.

[7] D. Deshwal, P. Sangwan, and D. Kumar, 'Feature Extraction Methods in Language Identification: A Survey', *Wirel. Pers. Commun.*, vol. 107, no. 4, pp. 2071–2103, Aug. 2019, doi: 10.1007/s11277-019-06373-3.

[8] 'Handbook of digital signal processing: engineering applications', *Choice Rev. Online*, vol. 26, no. 02, pp. 26-0940-26–0940, Oct. 1988, doi: 10.5860/CHOICE.26-0940.

[9] T. Mustofa, 'Implementation Speech Recognition for Robot Control Using MFCC and ANFIS', *J. Telemat. Inform.*, vol. 5, no. 2, Art. no. 2, Sep. 2017, doi: 10.12928/jti.v5i2.

[10] H. Heriyanto, S. Hartati, and A. E. Putra, 'Ekstraksi Ciri Mel Frequency Cepstral Coefficient (Mfcc) Dan Rerata Coefficient Untuk Pengecekan Bacaan Al-Qur'an', *Telematika*, vol. 15, no. 2, p. 99, Oct. 2018, doi: 10.31315/telematika.v15i2.3123.

[11] R. Hidayat, A. Bejo, and I. Muchlizar, 'SISTEM PENGENALAN SUARA SECARA REAL-TIME MENGGUNAKAN PEMROGRAMAN C', p. 7, 2019.

[12] M. Akhtar, 'Text Independent Biometric Authentication System Based On Voice Recognition', *Biom. Authentication*, p. 60, 2017.

Biodata Penulis

Ir.Ratna Hartayu,MT, lahir di Lumajang pada tahun 1965. Penulis Pertama memperoleh gelar Ir. Jurusan Teknik Fisika di Institut Teknologi Sepuluh Nopember Surabaya pada Tahun 1989. Kemudian melanjutkan pendidikan S2 jurusan Teknik Elektro di Institut Teknologi Sepuluh Nopember Surabaya dan lulus Tahun 1999. Konsentrasi penelitian penulis pertama yaitu Sistem Pengaturan dan Kewirausahaan. Saat ini penulis adalah salah satu dosen di Jurusan Teknik Elektro pada Universitas 17 Agustus 1945 Surabaya.

Santoso ST.,MT, lahir di Jembrana-Bali pada tahun 1973. Penulis kedua memperoleh gelar ST Jurusan Teknik Elektronika di Institut Teknologi Nasional Malang pada Tahun 1997. Kemudian melanjutkan pendidikan S2 di Institut Teknologi Sepuluh Nopember Surabaya dan lulus Tahun 2012. Konsentrasi penelitian penulis kedua yaitu bidang Artificial Intellience dan Robotika. Saat ini

penulis adalah salah satu dosen di Jurusan Teknik Elektro pada Universitas 17 Agustus 1945 Surabaya.

Choirul Anam, lahir pada 12 agustus di Magetan - Jawa Timur pada tahun 2000. saat ini penulis ketiga merupakan mahasiswa aktif semester 7 Teknik Elektro di Universitas 17 Agustus 1945 Surabaya

Dimas Abdul Azis, lahir di Sidoarjo-Jawa Timur pada 11 November 2000. Penulis keempat merupakan mahasiswa aktif semester 7 Teknik Elektro di Universitas 17 Agustus 1945 Surabaya