

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi, terutama dalam bidang komputer, saat ini melaju dengan sangat cepat dan pesat. Kemajuan teknologi yang seiring dengan perkembangan ilmu pengetahuan ini beriringan dengan berbagai inovasi yang dihasilkan oleh teknologi pada zamannya. Inovasi-inovasi tersebut memberikan manfaat positif bagi kehidupan manusia ke depannya, termasuk dalam berbagai aspek kehidupan, seperti pemahaman dan pemanfaatan suara.

Menurut (Prayoga, A, and Waluyo, 2019), suara merupakan fenomena fisik yang dihasilkan oleh getaran suatu benda yang berupa sinyal analog dengan amplitudo yang berubah secara kontinu terhadap waktu. Suara atau bunyi biasanya merambat melalui udara dan tidak bisa merambat melalui ruang hampa. Getaran ini mempunyai pola gelombang tertentu yang disebut *waveform* (bentuk gelombang). Suara juga merupakan bagian dari manusia; suara yang dimiliki oleh manusia memiliki keragaman bentuk yang berbeda-beda serta memiliki karakteristik yang dapat dibedakan berdasarkan gender atau identitas, baik laki-laki maupun perempuan.

Salah satu cara dalam proses identifikasi seseorang adalah dengan mengidentifikasi ciri biometriknya (identitas manusia). Teknologi biometrik merupakan suatu teknik pengenalan diri menggunakan bagian tubuh atau perilaku manusia. Biometrik merupakan salah satu cara yang paling praktis untuk mengidentifikasi individu manusia dengan hasil yang baik dan cepat melalui

karakteristik biologis yang unik, di mana setiap individu manusia itu berbeda. Salah satu sistem yang ada pada teknologi biometrik adalah pengenalan suara (Suhendra, Parlaungan, and Sumardi, 2023).

Pengenalan suara atau *voice recognition* merupakan bagian dari sub-teknologi biometrik, dengan sistem teknologi kecerdasan buatan (AI) yang memanfaatkan fitur unik dari suara manusia untuk menciptakan *voiceprints* atau sidik suara. Sama halnya dengan sidik jari, sidik suara juga mempunyai keunikan yang hanya dimiliki oleh satu individu (Nautsch *et al.*, 2019). Di dalam proses ekstraksi fitur terdapat banyak cara, salah satunya menggunakan *Mel-Frequency Cepstral Coefficients* (MFCC) yang menangkap karakteristik biometrik suara yang penting.

Mel-Frequency Cepstral Coefficients (MFCC) adalah fitur yang diekstraksi dari sinyal suara yang digunakan untuk representasi audio dalam bentuk yang lebih kompak dan informatif. MFCC menangkap karakteristik penting dari sinyal suara, terutama yang relevan dengan persepsi pendengaran manusia (Sasilo, Saputra, and Ningrum, 2022). Salah satu alasan menggunakan MFCC adalah kemampuannya untuk mendekati cara kerja telinga manusia lebih baik dibandingkan metode lainnya. Pendekatan ini dapat dicapai karena MFCC menggunakan filter bank dengan filter yang disusun secara linear pada frekuensi rendah dan logaritmik pada frekuensi tinggi. Dengan demikian, MFCC lebih sensitif terhadap sinyal frekuensi rendah (Shoalihin and Rakun, 2020). Setelah fitur-fitur sinyal suara diekstraksi, sinyal tersebut dapat diklasifikasikan menggunakan berbagai algoritma, salah satunya adalah *Convolutional Neural Networks* (CNN), yang dikenal efektif dalam tugas klasifikasi.

Convolutional Neural Networks (CNN) telah terbukti sangat efektif dalam tugas klasifikasi yang melibatkan data sensorik berdimensi tinggi, seperti gambar dan audio. Arsitekturnya sangat cocok untuk menangkap hierarki spasial dalam data, yang bermanfaat untuk memahami sinyal audio yang kompleks (Zaman *et al.*, 2023). CNN dapat diadaptasi untuk berbagai jenis tugas klasifikasi audio, termasuk klasifikasi genre musik, pengenalan suara, dan klasifikasi suara lingkungan, menjadikannya pilihan serbaguna bagi para peneliti dan praktisi (Kim, Lee, and Nam, 2019). CNN juga efektif untuk memproses data *grid-like* seperti gambar dan sinyal suara (Arafah, 2022).

Berdasarkan uraian latar belakang, penelitian ini dilakukan dengan judul "Implementasi MFCC dan CNN pada *Gender Voice Recognition*" dengan ekstraksi fitur menggunakan MFCC dan klasifikasi suara berdasarkan gender atau identitas antara laki-laki dan perempuan menggunakan algoritma CNN.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, dapat dijabarkan rumusan masalah penelitian adalah sebagai berikut:

- a. bagaimana membuat sistem *Gender Voice Recognition* yang efektif dan akurat?
- b. bagaimana menerapkan metode *Mel-Frequency Cepstral Coefficients* (MFCC) untuk ekstraksi fitur suara?
- c. bagaimana mengaplikasikan algoritma *Convolutional Neural Networks* (CNN) dalam proses klasifikasi suara berdasarkan gender?

1.3 Ruang Lingkup Penelitian

Berdasar dengan masalah yang diangkat, ruang lingkup dari penelitian ini adalah:

- a. sistem mengukur akurasi klasifikasi suara berdasarkan gender.
- b. sistem mampu mengklasifikasi gender berdasarkan suara yang dimasukkan.
- c. suara diekstrak menggunakan teknik *Mel-Frequency Cepstral Coefficients* (MFCC).
- d. suara yang diekstrak diidentifikasi menggunakan algoritma *Convolutional Neural Networks* (CNN).
- e. *dataset* yang digunakan adalah *VoxCeleb*.
- f. *dataset* berjumlah 5.993 data suara.
- g. *dataset* dipisahkan berdasarkan gender: *male* (laki-laki) dan *female* (perempuan).
- h. *dataset male* sebesar 3.682 data suara & *female* sebesar 2.311 data suara

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk mengetahui tingkat akurasi dari metode yang digunakan dalam *Gender Voice Recognition*, yaitu dengan MFCC dan CNN. Secara khusus, tujuan dari penelitian ini adalah:

- a. mengimplementasikan MFCC dalam proses ekstraksi fitur suara.
- b. mengimplementasikan algoritma CNN dalam proses klasifikasi suara berdasarkan gender.

- c. mengukur tingkat akurasi CNN dalam mengklasifikasikan suara antara *male* dan *female*.

1.5 Manfaat Penelitian

Adanya penelitian ini, diharapkan dapat memberi manfaat bagi lingkungan mahasiswa di kampus maupun umum, di antaranya:

- a. sebagai dasar dalam penelitian pengembangan sistem *Gender Voice Recognition* terutama di wilayah kampus UTDI (Universitas Teknologi Digital Indonesia).
- b. mengeksplorasi bagaimana MFCC dapat digunakan untuk ekstraksi fitur suara.
- c. mengeksplorasi bagaimana CNN dapat diimplementasikan untuk mengklasifikasikan fitur suara ke dalam kategori gender yang sesuai.
- d. mengetahui akurasi dari metode yang digunakan.