

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi memberikan kemudahan berkomunikasi setiap orang. Hal ini berjalan seiring dengan perkembangan komputer, telepon, dan alat komunikasi lainnya. Kondisi ini menyebabkan komunikasi menjadi hal yang penting dalam menjalin hubungan sosial [5].

Menanggapi segala masalah yang terjadi, aktivitas mengubah suara menjadi teks perlu dilakukan. Hal tersebut sangat berguna bagi penyintas tuna wicara. Oleh karena itu, aplikasi mengubah suara menjadi teks diperlukan untuk mempermudah tuna wicara untuk berbicara. Aplikasi tersebut diharapkan mempermudah pengguna dalam mengubah suara menjadi teks secara otomatis [5]. Dengan demikian, komunikasi yang baik bisa dilakukan kapan pun dan di mana pun.

Kemampuan berbicara yang dimiliki oleh setiap orang berbeda-beda. Hal ini berkaitan dengan tumbuh kembang seseorang, mulai dari anak. Tumbuh kembang anak ada yang cepat dan ada yang lambat. Tumbuh kembang anak yang lambat bisa berdampak bagi anak dalam melatih kemampuan berbicara, berkomunikasi, berbahasa, dan melakukan interaksi. Kondisi ini sangat mengkhawatirkan bagi orang tua karena bisa menyebabkan timbulnya autisme bagi anak [6, 7].

Masalah keterlambatan anak dalam berbicara banyak memunculkan rasa khawatir terhadap orang tua. Berdasarkan survei penelitian yang dilakukan oleh Hidajati tahun 2007, terdapat kira-kira 22,9% anak mengalami gangguan dalam berbicara dan berbahasa [6]. Kondisi tersebut sangat mengkhawatirkan orang tua karena dapat memengaruhi kemampuan anak dalam mengikuti proses pembelajaran di usia sekolah. Kondisi tersebut perlu menjadi perhatian dan tanggung jawab orang tua [8]. Tanggung jawab tersebut berupa memberikan intervensi pada anak yang memiliki kebutuhan khusus [7] dan memberi perangsang bagi anak sejak usia satu sampai dua tahun [6].

Pada konteks melatih kemampuan berbicara, orang tua harus menjadi tempat pertama bagi anak [6, 9]. Orang tua diberi kewajiban dalam mendidik dan membimbing anaknya. Hal tersebut dimulai dari pengenalan suara [9], pengenalan bahasa [9], terapi wicara [7], dan terapi perilaku. Menyikapi hal tersebut, kemampuan berbicara akan dibahas lebih dalam pada konteks kemampuan

berbicara ke teks.

Berbicara adalah hal umum yang dilakukan oleh manusia sejak usia dini. Berbicara merupakan suatu keterampilan interaksi dalam kehidupan sehari-hari. Interaksi adalah dasar dari proses sosial yang dilakukan oleh manusia untuk bertukar informasi dengan orang lain untuk bertahan hidup. Interaksi ini memiliki maksud untuk menjalin hubungan yang dinamis antar individu, antar kelompok, atau pun individu dengan kelompok [10].

Berbicara bisa dilakukan dalam bahasa apa pun. Pada penelitian ini, bahasa yang digunakan adalah Bahasa Indonesia. Bahasa tersebut akan digunakan untuk mengirimkan pesan ke teks. Pesan tersebut berupa pesan suara yang berjumlah satu sampai dua kata dengan jumlah dua sampai enam suku kata.

Pada penelitian ini, ada tiga aplikasi akan digunakan dalam membuat pesan suara. Aplikasi tersebut ialah PyCharm, Windows Media Player, dan Visual Studio Code. Aplikasi PyCharm akan menerjemahkan semua pembicaraan yang diucapkan oleh manusia dengan menggunakan bahasa pemrograman Python yang handal [11]. Aplikasi ini juga akan digunakan pengujian data suara dengan penerapan algoritma *Convolutional Neural Network* (CNN). Algoritma CNN tersebut menguji keakuratan data suara dengan menggunakan model pra-latihan [12]. Aplikasi Windows Media Player digunakan untuk memutar berkas audio. Selain itu, aplikasi Visual Studio Code digunakan dalam proses pengujian model.

Menerjemahkan pesan ke teks dilakukan dengan menggunakan perintah audio. Perintah audio berfungsi untuk menerjemahkan pesan suara yang diucapkan oleh seseorang. Menerjemahkan pesan suara dilakukan dengan cara seseorang berbicara satu sampai dua kata. Kata yang dibicarakan oleh seseorang akan muncul pada tampilan aplikasi komputer. Jika kata yang keluar sesuai dengan hasil yang diharapkan, maka pesan suara akan berhasil. Namun, pesan suara akan gagal dan akan ada peringatan **"Maaf suara Anda kurang jelas"** jika hasil yang keluar tidak sesuai dengan pesan suara yang diucapkan melalui pesan audio [13]. Selain itu, pesan suara juga bisa gagal jika seseorang tidak berbicara selama lebih dari 0,8 detik sejak program tersebut dijalankan [14].

Pada penelitian kali ini, sampel data suara diperlukan sebagai pedoman utama. Sampel tersebut berupa himpunan data. Himpunan data yang digunakan pada penelitian ini adalah himpunan data sendiri dan himpunan data dari internet. Himpunan data sendiri akan dibuat menggunakan bahasa pemrograman Python yang sangat handal [11] dan dijalankan pada terminal. Data yang sesuai akan disimpan ke dalam berkas audio WAV pada aplikasi komputer. Berkas audio

tersebut akan diekspor ke himpunan data Kaggle. Sedangkan, himpunan data dari internet akan dicari menggunakan aplikasi Kaggle.

Pada bagian akhir, penelitian ini akan dilakukan publikasi. Publikasi ini dilakukan agar aplikasi melatih kemampuan berbicara tidak hanya digunakan untuk diri sendiri. Aplikasi ini juga bisa digunakan oleh anak usia dini, semua pria, dan semua wanita dalam melatih kemampuan berbicara kata demi kata [15].

1.2 Rumusan Masalah

Pada bagian ini disampaikan rumusan masalah yang diperbincangkan, yaitu:

- Bagaimana penerapan algoritma *Convolutional Neural Network* (CNN) untuk konversi audio ke teks?
- Bagaimana mengukur kinerja dari model yang digunakan pada *Convolutional Neural Network* (CNN)?

1.3 Batasan Permasalahan

Pada bagian ini disampaikan batasan masalah yang hendak dicapai, yaitu:

- Jumlah kata yang diucapkan oleh seseorang dibatasi yaitu satu sampai dua kata dengan jumlah dua sampai enam suku kata.
- Algoritma yang digunakan untuk pengujian model adalah *Convolutional Neural Network* (CNN).

1.4 Tujuan Penelitian

Pada bagian ini disampaikan tujuan yang hendak dicapai dari penelitian ini. Ada beberapa tujuan yang sesuai dengan rumusan masalah yang telah disampaikan sebelumnya.

- Untuk menguji hasil pembicaraan manusia dari audio menjadi teks melalui proses pengujian model dengan penerapan algoritma *Convolutional Neural Network* (CNN).
- Untuk mengukur tingkat akurasi melalui pengujian model yang terdapat pada algoritma *Convolutional Neural Network* (CNN) yang berupa akurasi, presisi, *recall*, nilai F1, rata-rata makro, dan rata-rata berbobot.

1.5 Manfaat Penelitian

Pada bagian ini diuraikan manfaat yang akan diperoleh dari penelitian ini. Ada beberapa manfaat yang disampaikan haruslah sesuai dengan lingkup kerja penelitian.

- Memperoleh hasil dari konversi model pembicaraan manusia dari audio ke teks dengan audio yang memiliki format WAV.
- Memperoleh tingkat akurasi, tingkat kerugian, matriks *confusion*, dan laporan klasifikasi dari kinerja model pada Python.

1.6 Sistematika Penulisan

Berisikan uraian singkat mengenai struktur isi penulisan laporan penelitian, dimulai dari Pendahuluan hingga Simpulan dan Saran.

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN
Bab pertama merupakan bagian awal dari sebuah penelitian. Bagian ini berisi tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, sistematika penulisan skripsi, dan spesifikasi sistem. Isi dari semua komponen pada halaman pendahuluan berisi tentang hal-hal yang ingin dibahas pada penelitian tersebut.
- Bab 2 LANDASAN TEORI
Bab kedua merupakan bagian yang menjelaskan teori yang dibutuhkan pada suatu penelitian. Teori tersebut mengenai kemampuan berbicara, wicara, penerapan algoritma CNN, penggunaan tensorflow, audio, *One Hot Encoder*, dan model h5. Teori ini bertujuan agar seseorang bisa mengirimkan pesan ke dalam suatu aplikasi komputer melalui kemampuan berbicara manusia menggunakan kode program pada aplikasi komputer. Aplikasi komputer yang akan digunakan adalah PyCharm, Visual Studio Code, Kaggle, dan Windows Media Player.
- Bab 3 METODOLOGI PENELITIAN
Bab ketiga merupakan bab yang berisi metode dalam proses penelitian. Metode ini ialah pra-proses untuk memperoleh hasil yang diharapkan. Ada 12 metode yang digunakan, yaitu studi pustaka, metode pengumpulan data,

metode pengambilan sampel, metode perancangan, metode evaluasi, metode dokumentasi, model penggunaan aplikasi, metode penggunaan himpunan data, metode implementasi himpunan data, proses pembuatan model, proses menyimpan model, dan proses melihat model.

- Bab 4 HASIL DAN DISKUSI

Bab ini merupakan bab yang berisi tentang hasil yang dicapai selama kegiatan penelitian ini berlangsung. Hasil yang dicapai berupa pengujian untuk mengucapkan satu atau dua kata. Kata tersebut bisa merupakan kata dalam Bahasa Indonesia. Pengujian tersebut dilakukan selama dua sampai tiga bulan. Lama waktu pengujian adalah satu sampai tiga detik dalam satu sesi pengujian.

Hasil dari kegiatan penelitian yang dilaksanakan juga merupakan hasil dari implementasi data dari metode pengumpulan data melalui proses merekam data suara pada aplikasi PyCharm. Data suara tersebut akan disimpan dan dilakukan konversi ke WAV.

Data suara yang telah dilakukan proses konversi ke WAV, dilakukan pengujian model. Pengujian model tersebut dilakukan dengan menggunakan data sendiri. Pengujian dilakukan dengan menggunakan model pra-pelatihan pada Kaggle dan proses menjalankan model melalui Visual Studio Code. Hasilnya akan menampilkan hasil dari model yang berupa akurasi data, tingkat kerugian, *confusion* matriks, dan laporan klasifikasi. Model tersebut akan disimpan dalam bentuk *High-Density Fiberboard* (HDF) dengan format h5.

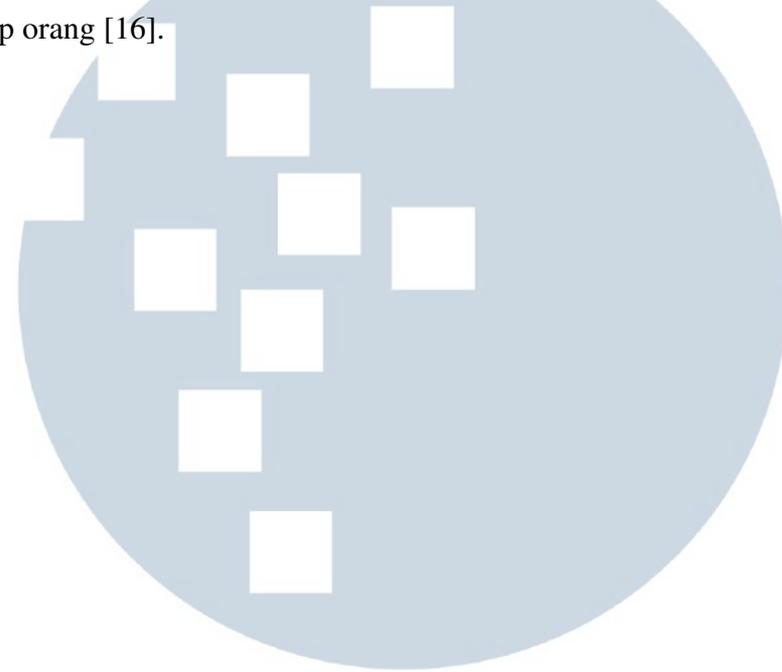
- Bab 5 KESIMPULAN DAN SARAN

Bab ini merupakan bab yang berisi hasil akhir dari penelitian yang telah dilaksanakan. Hasil yang telah dicapai akan disimpulkan secara menyeluruh. Hasil yang mengalami kekurangan atau tidak sesuai dengan yang diharapkan akan diberikan saran dan rencana pengembangan yang lebih lanjut.

1.7 Spesifikasi Sistem

Spesifikasi sistem yang akan digunakan untuk mengerjakan penelitian kali ini adalah perangkat lunak dan perangkat keras. Perangkat lunak yang digunakan berupa Windows, Linux, Google Chrome, dan Overleaf. Perangkat keras yang digunakan berupa Mouse, Keyboard, mikrofon, dan CPU.

Perangkat lunak dan perangkat keras merupakan komponen yang terpenting dalam spesifikasi sistem. Tanpa kedua perangkat tersebut, maka penelitian ini tidak dapat berjalan sesuai rencana. Tujuannya adalah untuk memperluas studi sebelumnya dengan mengevaluasi penggunaan perangkat lunak melalui pengenalan suara setiap orang [16].



UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA