

BAB I

PENDAHULUAN

1.1 Latar Belakang

Mengutip KBBI, literasi merupakan kemampuan seseorang membaca dan menulis. Kemampuan literasi, khususnya membaca, adalah aspek penting dalam perkembangan individu. Namun, di Indonesia sendiri tingkat literasi baca masih tergolong rendah. Berdasarkan data yang dipublikasikan oleh PISA (*Programme for International Student Assessment*) pada website resmi OECD (*Organization for Economic Co-operation and Development*) skor literasi membaca Indonesia hanya mencapai angka 359, jauh dibawah rata-rata yang mencapai skor 475 [1]. Fakta tersebut didukung dengan penelitian UNESCO yang menyatakan bahwa hanya 0,001% dari masyarakat Indonesia yang memiliki minat baca [2].

Hal-hal tersebut diperparah dengan pola konsumsi informasi masyarakat Indonesia. Sebuah survei mendapatkan data bahwa 91% pengguna YouTube di Indonesia lebih memilih konten video pendek [3]. Banyaknya peminat video pendek ini sejalan dengan rendahnya minat membaca [4]. Dengan video pendek, sebuah hal atau informasi akan didapatkan dengan sangat singkat, berbanding terbalik dengan membaca teks panjang atau buku.

Kondisi ini cukup berdampak pada berbagai kalangan terutama kalangan pekerja dan anak-anak. Cukup banyak pekerja yang terlalu sibuk hingga sulit menyempatkan waktu membaca, padahal tidak sedikit yang sebenarnya masih berminat untuk membaca. Efek domino terjadi pada anak-anak mereka di mana kebiasaan bercerita sebelum tidur ditinggalkan karena kesibukan. Padahal dengan cerita sebelum tidur atau *bedtime story* sangat membantu anak untuk meningkatkan literasi dini dan membangun minat untuk membaca [5].

Audiobook hadir sebagai solusi demi mempermudah akses literasi karena berbentuk audio. Walaupun berbentuk audio, tetapi esensi dari membaca sendiri tidak hilang karena konten yang disajikan dan informasi yang didapat sama dengan bentuk teks [6]. *Audiobook* juga tidak jauh berbeda dengan cerita seperti dongeng yang diceritakan turun temurun dari mulut ke mulut yang mana esensi dari cerita

tersebut tetap sama [7]. Namun, *audiobook* yang tersedia masih terbatas pada buku-buku tertentu, terutama dalam bahasa Indonesia yang jumlahnya lebih sedikit. Teknologi *Text-to-Speech* menawarkan alternatif di mana lebih fleksibel karena memungkinkan konversi tulisan menjadi audio dalam berbagai bentuk sumber seperti buku, artikel, maupun dokumen.

Seperti yang dijelaskan sebelumnya bahwa *Text-to-Speech* ini adalah teknologi yang memungkinkan perubahan teks menjadi audio [8]. TTS merupakan cabang dari NLP atau *Natural Language Processing*. Pendekatan dalam pengembangan model TTS sendiri cukup beragam mulai dari metode konvensional seperti *Rule Based* hingga metode *Deep Learning* seperti Tacotron (berbasis RNN) dan FastSpeech2 (berbasis Transformer).

FastSpeech2 yang berbasis Transformer menawarkan kualitas sintesis suara yang lebih baik dibanding metode lainnya. Sayangnya referensi metode FastSpeech2 ini cukup banyak yang tidak terbuka secara publik aksesnya. Secara keseluruhan pun jumlah pengembangan TTS sendiri bisa dibilang masih sedikit dibanding cabang NLP lainnya. Belum lagi tidak adanya dataset publik berbahasa Indonesia yang khusus untuk bidang TTS. Maka untuk pengembangan model dengan bahasa Indonesia menjadi tantangan tersendiri dalam penelitian ini.

Oleh karena itu, penelitian dengan judul `PENGEMBANGAN MODEL TEXT-TO-SPEECH BAHASA INDONESIA MENGGUNAKAN METODE FASTSPEECH2` dibuat dengan tujuan mengembangkan model TTS Bahasa Indonesia berbasis FastSpeech2. Dataset yang digunakan nantinya akan dibuat khusus mengikuti format dataset LJSpeech yang biasa digunakan untuk model FastSpeech2 karena keterbatasan dataset publik untuk *speech* berbahasa Indonesia. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan teknologi TTS bahasa Indonesia serta menjadi solusi alternatif dalam upaya meningkatkan literasi masyarakat Indonesia khususnya untuk kalangan pekerja dan anak-anak.

1.2 Rumusan Masalah

Terdapat beberapa pertanyaan utama yang harus dijawab dalam penelitian ini untuk mendapatkan solusi atas permasalahan yang ada. Rumusan masalah yang akan dibahas meliputi :

- a. Bagaimana proses pembuatan dataset yang sesuai dengan format LJSpeech untuk melatih model TTS bahasa Indonesia?
- b. Bagaimana cara mengembangkan model *Text-to-Speech* (TTS) bahasa Indonesia menggunakan FastSpeech2?
- c. Bagaimana cara mengukur kualitas hasil sintesis suara dari model yang dikembangkan?

1.3 Batasan Masalah

Penelitian ini dibatasi sampai pengembangan model dan evaluasi hasil sintesis suara dalam bahasa Indonesia. Adapun batasan masalah lainnya sebagai berikut :

- a. Bahasa
 - Penelitian ini terbatas pada Bahasa Indonesia untuk dari dataset hingga hasil sintesis suara dari model.
- b. Evaluasi Model
 - Penilaian kinerja model dilakukan dengan mengukur kualitas hasil sintesis suara berdasarkan *Mel Cepstral Distortion* dan *Mean Opinion Score*.
- c. Batasan Teks Input
 - Teks yang dapat disintesis oleh model dibatasi pada teks input langsung dan teks file seperti format pdf.
- d. Ruang Lingkup Penggunaan
 - Model akan diintegrasikan dengan sebuah aplikasi sederhana berbasis web pada tahap akhir penelitian.

1.4 Tujuan

Berdasarkan rumusan masalah yang telah dijelaskan sebelumnya, ditetapkan beberapa tujuan penelitian ini sebagai berikut :

- a. Mengetahui cara membuat dataset dengan format LJSpeech untuk melatih model TTS.
- b. Mengetahui cara mengembangkan model TTS bahasa Indonesia menggunakan FastSpeech2.
- c. Mengetahui cara mengukur kualitas sintesis suara dari model yang dikembangkan.

1.5 Manfaat

Diharapkan dengan berjalannya penelitian ini akan memberikan manfaat yang positif. Manfaat yang dapat dihasilkan dari penelitian ini terdapat beberapa bagian, yaitu :

- a. Peneliti

Memahami dan mendapatkan keterampilan dalam pengembangan model *Text-to-Speech* berbahasa Indonesia. Kemudian menambah pengalaman dalam merancang, membangun, dan mengevaluasi model TTS bahasa Indonesia.

- b. Keilmuan Informatika

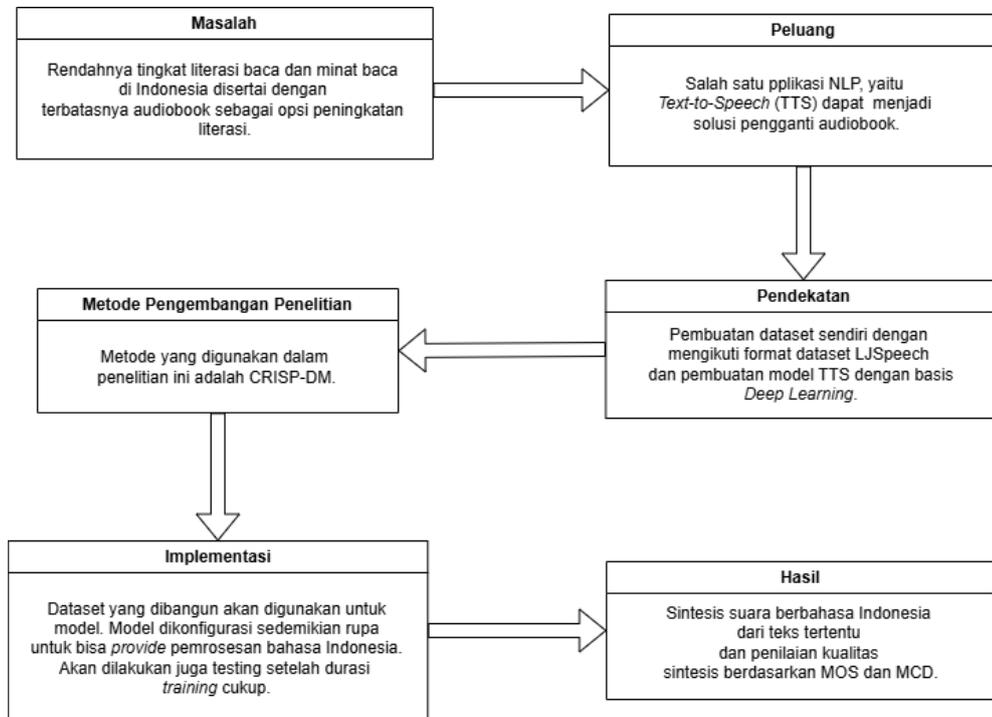
Berkontribusi pada literatur terkait *Text-to-Speech* berbahasa Indonesia, khususnya penggunaan metode FastSpeech2 dalam TTS bahasa Indonesia sendiri yang masih tergolong sangat jarang dilakukan penelitiannya.

- c. Masyarakat

Memberikan alternatif bahan literasi berbentuk audio dalam upaya meningkatkan literasi Indonesia dan juga sebagai alternatif *audiobook* yang cukup terbatas jumlahnya dalam bahasa Indonesia.

1.6 Kerangka Pemikiran

Untuk menunjang penelitian ini, maka dibentuklah sebuah kerangka pemikiran seperti yang divisualisasikan Gambar 1.1.



Gambar 1. 1 Kerangka Pemikiran