

## BAB 5 KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, diperoleh kesimpulan bahwa sistem aplikasi transkripsi percakapan telah berhasil dikembangkan dengan baik menggunakan model *Automatic Speech Recognition (ASR) Wav2Vec2*. Performa model dengan perhitungan metrik *Word Error Rate (WER)* menunjukkan hasil yang sangat baik hanya dengan menggunakan 6 jam dataset suara Bahasa Indonesia. Sedangkan waktu transkripsi cukup efisien dengan rata-rata persentase waktu transkripsi adalah 16,98% dari durasi audio input.

Untuk mendapatkan hasil akurasi model *ASR* yang baik dan optimal, dilakukan tiga skenario pengujian utama. Skenario pertama dilakukan dengan mencari *hyperparameter* yang berpengaruh dan nilai optimalnya menggunakan metode *manual search* guna mempercepat proses pengujian. Dari hasil pengujian, diperoleh 2 *hyperparameter* yang sangat berpengaruh pada performa model yaitu *learning rate* dan *epochs* yang masing-masing mempunyai nilainya optimalnya adalah 0,0003 untuk *learning rate* dan 60 untuk *epochs*.

Skenario kedua dilakukan dengan menguji pengaruh penambahan *N-Gram Language Models* ke dalam proses model *Wav2Vec2* yang sudah dilatih. Pengujian menunjukkan bahwa terjadi peningkatan performa yang sangat signifikan setelah melakukan pengujian dengan skenario menggunakan variasi nilai *N*. Hasil optimal ditunjukkan oleh *N* dengan nilai 2 dengan peningkatan performa *WER* sebesar 54% dari *WER* semula 0,283 menjadi 0,129.

Skenario terakhir adalah membandingkan model *Wav2Vec2* satu bahasa dengan *Wav2Vec2* versi multi bahasa (*XLSR-Wav2Vec2*) dengan semua parameter yang sama setelah melalui skenario pengujian sebelumnya. Dari hasil pengujian disimpulkan bahwa model *Wav2Vec2* multi bahasa menunjukkan performa yang lebih baik dengan menggunakan *N-Gram* maupun tanpa *N-Gram*. Hasil metrik pengujian model *Wav2Vec2* multi bahasa tanpa *N-Gram* menunjukkan hasil 0,216 *WER* yang menunjukkan adanya peningkatan performa dengan penurunan *WER* sebesar 23% dari model *Wav2Vec2* satu bahasa. Sedangkan model *Wav2Vec2* multi bahasa dengan *N-Gram* menunjukkan hasil 0,104 *WER* dengan penurunan *WER* sebesar 17% dari model *Wav2Vec2* satu bahasa.

Hasil akhir model Wav2Vec2 yang telah dikembangkan dengan nilai *WER* 0,104 kemudian dilakukan pengujian evaluasi metrik lebih lanjut. Digunakan 3 evaluasi metrik yaitu *Precision*, *Recall*, dan *F1-Score* untuk mengukur performa model Wav2Vec2. Dari hasil evaluasi, didapat hasil yang baik dimana nilai *Precision* adalah 0,960; nilai *Recall* adalah 0,938; dan nilai *F1-Score* adalah 0,946 yang menunjukkan model ASR Wav2Vec2 yang dikembangkan baik dan stabil.

## 5.2 Saran

Saran yang dapat diberikan untuk penelitian selanjutnya adalah sebagai berikut.

1. Penentuan Hyperparameter yang digunakan selama masa pelatihan model sangat mempengaruhi performa model yang dihasilkan. Berdasarkan penelitian, sedikit saja perubahan bahkan dapat mengakibatkan model menjadi *overfitting* ataupun *underfitting*. Saran ini merupakan salah satu yang berpotensi dapat ditingkatkan dan diteliti lebih lanjut baik dengan mencoba lebih banyak skenario ataupun menggunakan *hyperparameter* lainnya.
2. Penambahan *Language Models* sangat mempengaruhi performa pengenalan suara. Berdasarkan penelitian, penambahan N-Gram *Language Models* mempengaruhi hingga 55% performa *Word Error Rate*, saran ini merupakan salah satu yang berpotensi dapat ditingkatkan dan diteliti lebih lanjut baik menggunakan korpus dengan ukuran yang lebih banyak ataupun menggunakan pendekatan *Language Models* lainnya.
3. Jumlah dan kualitas *dataset* suara berpotensi meningkatkan performa model *Automatic Speech Recognition* yang dibangun. Penelitian menggunakan dataset kurang lebih 6 jam suara bahasa Indonesia guna mempercepat proses *training model*, sehingga masih terdapat ruang yang sangat luas untuk dilakukan penelitian lebih lanjut terkait pengaruh dataset baik dari segi jumlah, durasi, kosa kata, dan lain lain.
4. Sistem transkripsi percakapan akan lebih efektif jika dapat mendeteksi tanda baca. Saran ini didapat ketika dilakukan evaluasi dengan pakar. Saran ini masih sangat berpotensi untuk diteliti lebih lanjut guna meningkatkan efektivitas sistem transkripsi percakapan.