

## BAB V KESIMPULAN DAN SARAN

### 5.1 Kesimpulan

Dengan menggunakan kombinasi *Dynamic Time Warping* (DTW) dan *Web Speech API*, penelitian ini berhasil mengembangkan aplikasi *speech-to-text*. Aplikasi ini dapat mengubah ucapan menjadi teks secara realtime. Namun, gangguan suara, seperti suara hujan, yang dapat memengaruhi akurasi transkripsi, yang akhirnya memengaruhi hasil penggunaan aplikasi. Hasil pengujian menunjukkan bahwa aplikasi dapat menghasilkan transkripsi dengan kesalahan lebih sedikit, yaitu dengan nilai kesalahan kata atau *Word Error Rate* (WER) terendah sebesar 6,8%. Namun, perlu diingat bahwa lingkungan rekaman suara dapat memengaruhi kinerja aplikasi yang membuat hasil akurasi cenderung berubah-ubah seperti terlihat pada nilai kesalahan kata atau *Word Error Rate* (WER) terbesar yaitu 15,4% sehingga diperlukan upaya untuk mengurangi gangguan suara untuk meningkatkan kualitas transkripsi.

Terkait dengan hasil jumlah kata yang dihapus lebih banyak dibandingkan dengan kata yang ditambahkan dan disubstitusi selama proses transkripsi, menunjukkan bahwa sistem cenderung menghasilkan transkripsi yang lebih singkat dibandingkan dengan teks referensi yang berasal dari hasil transkripsi manual. Hal ini dapat disebabkan oleh berbagai alasan, seperti kesulitan menemukan kata-kata tertentu, kecenderungan untuk menghilangkan kata-kata yang dianggap tidak penting, atau kurangnya pemahaman tentang konteks teks. Analisis menyeluruh diperlukan untuk meningkatkan keakuratan dan keandalan transkripsi serta penyesuaian atau pengembangan lanjutan dari implementasi algoritma pada sistem untuk meningkatkan kinerja aplikasi.

Hasil percobaan juga menunjukkan bahwa variasi dalam pola ucapan yang diucapkan dalam video dan pemahaman homofon (kata-kata yang terdengar sama tetapi memiliki makna yang berbeda) setiap kata pada ucapan pembaca dapat mempengaruhi hasil transkripsi. Hal ini menunjukkan betapa pentingnya mengembangkan sistem *speech-to-text* yang lebih kompleks dengan mempertimbangkan variasi ucapan dan pemahaman konteks. Oleh karena itu, dengan mempertimbangkan kesulitan yang terkait dengan lingkungan rekaman suara dan variasi dalam ucapan manusia, penelitian ini diharapkan dapat membantu dalam pengembangan aplikasi *speech to text* yang lebih akurat dan dapat diandalkan. Dengan memperoleh pemahaman yang lebih baik

tentang komponen yang memengaruhi kinerja aplikasi, pengembangan selanjutnya dapat diarahkan untuk meningkatkan kualitas dan kegunaan aplikasi dalam berbagai konteks.

## 5.2 Saran

Karena dalam penelitian ini masih terdapat kelemahan, salah satunya yaitu implementasi algoritma yang digunakan masih terlalu dasar, maka kinerja dan akurasi aplikasi dapat terbatas. Keterbatasan dalam hal ini adalah implementasi algoritma *Dynamic Time Warping* yang diterapkan hanya sebagai perhitungan perbandingan antara transkripsi otomatis dan transkripsi manual. Algoritma yang kurang canggih juga mungkin tidak dapat menangani variasi pola ucapan yang kompleks atau kondisi lingkungan yang berbeda dengan baik, sehingga menghasilkan transkripsi yang tidak akurat. Integrasi dengan teknologi NLP yang lebih lengkap lagi mungkin akan dapat membantu meningkatkan pemahaman konteks dan akurasi transkripsi, terutama dalam kasus dengan kata-kata homofon atau frase yang ambigu. Dengan menerapkan saran-saran ini, diharapkan aplikasi ini dapat terus berkembang dan memberikan kontribusi yang lebih besar dalam mendukung proses pengembangan transkripsi suara secara otomatis.

