

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil dan analisis, berikut adalah beberapa kesimpulan yang dihasilkan dari penelitian ini, yaitu:

1. Implementasi teknik autoencoder dan RNN pada studi kasus pembuat instrumen musik otomatis berdasarkan suara vokal manusia adalah dengan menyiapkan dua model *autoencoder* dan satu model RNN, di mana model *autoencoder* pertama digunakan untuk melatih data suara vokal dan model *autoencoder* kedua digunakan untuk melatih data suara instrumen musik. Hal ini dilakukan untuk memperoleh ekstraksi data yang disebut *latent* agar dapat dilatih pada model RNN. Model RNN menjadi model utama yang berisi pelatihan dari model *encoder autoencoder* pertama dan kemudian *output* dari model RNN akan dihubungkan ke bagian model *decoder autoencoder* kedua.
2. Arsitektur model autoencoder yang baik untuk mendapatkan *latent* dengan ukuran optimal dan ketika diprediksi menggunakan model RNN, *output* yang dihasilkan masih dapat didengar adalah dengan rekonstruksi *2-layers Conv1D*, *2-layers pooling*, dan *1-fully connected layer*. Model *encoder* memiliki susunan arsitektur *Conv1D(8x3) – AveragePooling(7) – Conv1D(2x3) – AveragePooling(7)*, sedangkan model *decoder* memiliki susunan arsitektur cermin dari model *encoder*, yaitu *UpSampling(7) – Conv1D(2x3) – UpSampling(7) – Conv1D(8x3)* dengan *fully connected layer Conv1D(1x3)*.
3. Jumlah layer yang dibutuhkan pada rekonstruksi model arsitektur model *autoencoder* dan model RNN untuk hasil evaluasi terbaik yaitu pada model *autoencoder* dengan rekonstruksi jumlah *3-layers* konvolusi 1 dimensi, *1-layer pooling*, dan *1-fully connected layer*.

Sedangkan pada model RNN dengan rekonstruksi jumlah 1-layer GRU dan 1-fully connected layer.

4. Teknik sisir (*combing*) dan DCT menjadi representasi yang baik untuk *preprocessing* data suara agar dapat diproses pada spesifikasi komputer rendah, karena diperoleh ukuran data *latent* yang kecil sehingga saat proses *training* model tidak mengkonsumsi memori yang banyak.
5. Hasil *test* dengar *output* dari skema *autoencoder* – RNN adalah suara terdengar jelas, namun tidak ada kreativitas di dalamnya karena masih seperti suara dari *input* vokal. Pada skema *autoencoder* – sisir (*combing*) – RNN adalah suara terdengar jelas tetapi tidak ada nada yang terdengar pada suara tersebut, sebaliknya, terdengar seperti tumpukan suara yang tidak beraturan dan ambigu yang disertai *noise*, hal ini menunjukkan model berusaha membuat kreativitas di dalamnya namun terbatas dikarenakan pada tahap *preprocessing* data yang belum tepat. Pada skema RNN – DCT adalah suara terdengar jelas dan bernada dengan cukup harmonis, namun sekali lagi hal ini dipengaruhi oleh teknik *preprocessing* data, semakin tepat teknik *preprocessing* data yang digunakan, maka semakin baik model yang didapat untuk menghasilkan *output* suara instrumen musik yang baik.

5.2 Saran

Meskipun penelitian ini memiliki beberapa keterbatasan dalam metodologi dan alat yang digunakan, namun penelitian ini memberikan kontribusi yang signifikan dalam bidang pemrosesan suara. Fakta-fakta baru yang ditemukan dalam penelitian ini dapat menjadi dasar bagi pengembangan penelitian di masa depan. Adanya potensi besar pada penelitian ini, memperkuat harapan bahwa dengan menggunakan teknik yang lebih baru, hasil penelitian akan semakin ditingkatkan. Oleh karena itu, para peneliti di masa depan dapat memanfaatkan hasil penelitian ini untuk meningkatkan dasar pengetahuan di bidang pemrosesan suara dan melakukan penelitian yang lebih akurat dan dapat diandalkan.