

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Menciptakan instrumen musik perlu adanya pengetahuan tentang musik, bahkan ketika seorang komposer ingin menciptakan sebuah instrumen musik untuk dijadikan sebagai *background* dari sebuah lagu, terdapat banyak tahap yang perlu dilakukan. Mulai dari menentukan tema, *genre*, keharmonisan nada, dan lirik dari lagu yang diciptakan. Tidak semua orang dapat membuat sebuah instrumen atau *background* musik dari sebuah lagu. Berdasarkan hasil survei terhadap 55 responden melalui *google form*, didapatkan hasil bahwa mayoritas atau 89,1% responden tidak memiliki kemampuan dalam menciptakan atau membuat musik instrumen dikarenakan kurangnya pengetahuan dan pengalaman tentang membuat nada pada instrumen musik. Sebaliknya, sebagian kecil dari 10,9% responden menyatakan bahwa mereka memiliki kemampuan untuk membuat musik instrumen dengan beberapa alasan yang berbeda, seperti memiliki basic musik, pernah belajar atau berlatih, serta ada kemauan untuk mencari cara dan belajar. Hasil survei ini menunjukkan bahwa tidak semua orang memiliki kemampuan atau keahlian dalam menciptakan musik instrumen, namun sebagian kecil dari mereka telah memahami dan menguasai dasar-dasar musik yang memungkinkan mereka untuk membuat instrumen musik. Sehingga berdasarkan masalah ini, bagaimana caranya agar semua orang dapat membuat instrumen musik otomatis, sesuai dengan nyanyian orang tersebut [1][2][3].

Musik adalah pola melodi, ritme, harmoni, tempo/dinamika, dan timbre yang digabungkan untuk menciptakan pengulangan, variasi, dan kontras [4][5]. Sedangkan lagu adalah suatu kombinasi nada yang berirama dan harmonis, yang dihasilkan melalui instrumen alat musik dan suara vokal manusia. Maka dari itu, musik menjadi komponen penting dari sebuah lagu agar menghasilkan keharmonisan antara instrumen dan lirik lagu tersebut [6][7][8].

Terdapat salah satu jenis musik yang unik, dimana musik yang dihasilkan tanpa melalui instrumen dari alat musik. Akapela, adalah jenis musik yang dilakukan oleh sekelompok penyanyi yang bernyanyi tanpa iringan instrumen musik. Dengan kata lain, suara penyanyi menghasilkan semua iringan atau menjadi pengganti suara alat musik [9][10].

Penelitian dengan topik seperti *machine learning* dan *deep learning* yang fokusnya di bidang teknologi, menjadikan musik sebagai objek yang menarik untuk diteliti [11][12][13][14][15]. Salah satunya adalah pembuatan musik otomatis berbasis *deep learning* menggunakan *Convolutional Neural Network* (CNN) dan *Recurrent Neural Network* (RNN) [16][17][18][19]. Algoritma CNN adalah salah satu algoritma *deep learning* yang dapat diaplikasikan untuk ekstraksi fitur pada kasus *voice to text* [20][21]. Sedangkan algoritma RNN adalah algoritma *deep learning* yang menerima *input* dan *output* berupa *time series*, di mana modelnya dapat mengingat urutan *inputan* karena di dalamnya memiliki memori internal [22][23]. Oleh karena itu, RNN seringkali dipergunakan untuk *input* data sekuensial [24][25].

Penelitian yang dilakukan oleh Shaily Malik, Gaurav Arora, Anvaya Ahlawat, dan Mandeep Payal dengan judul penelitian “Music Generation Using Deep Learning – Char RNN (2021)” yang mengambil topik tentang pengaplikasian algoritma RNN *Long Short-Term Memory* (LSTM) untuk membuat musik baru otomatis. Penelitian ini menggunakan *dataset* berbentuk notasi musik yang *ditrain* dengan berbagai arsitektur LSTM. Penelitian ini menunjukkan bahwa arsitektur LSTM dapat digunakan untuk membuat musik meskipun tidak sempurna dan juga memiliki potensi untuk menghasilkan musik dengan kompleks yang lebih tinggi. LSTM terbukti menjadi model yang baik untuk menangkap depedensi berskala panjang [26].

Selain itu, penelitian yang dilakukan oleh Jiatong Xie dengan judul penelitian “A Novel Method of Music Generation Based on Three Different Recurrent Neural Networks (2020)” di mana penelitian ini menggunakan jaringan saraf RNN seperti LSTM dan *Gated Recurrent Unit* (GRU) untuk menyusun inti jaringan dalam memprediksi nada dari suatu melodi dan menghasilkan musik

baru. Penelitian ini menggunakan *dataset* dalam format *Musical Instrument Digital Interface* (MIDI) karena dianggap mempunyai struktur data yang sederhana. Penelitian ini menunjukkan bahwa model RNN mampu merekam data yang bersifat *time series* dan memprediksi nada berdasarkan informasi sebelumnya. Model ini mencapai akurasi lebih dari 80% dari hasil percobaan dan tanpa mengalami *overfitting* [27].

Berdasarkan uraian diatas, penulis tertarik untuk melakukan penelitian tentang generator musik instrumental otomatis pada akapela menggunakan metode *deep learning*. *Deep learning* menjadi salah satu metode yang sering digunakan untuk studi kasus pengolahan suara, seperti memproses sinyal suara secara sekuensial dan mengekstrak fitur-fitur penting dari sinyal tersebut dapat menggunakan Conv1D dan untuk memodelkan hubungan sekuensial antara data masukan dengan data keluaran, yang sangat berguna dalam pengolahan data suara yang bersifat sekuensial dapat menggunakan RNN. Metode yang akan digunakan dalam penelitian ini mencakup tiga skema, yaitu skema *autoencoder – RNN*, skema *autoencoder–siris (combing)*, dan skema *Discrete Cosine Transform (DCT) – RNN*. Setiap skema akan dieksplorasi dengan beberapa model, dan kemudian akan dilakukan evaluasi untuk menentukan spesifikasi model terbaik pada setiap skema. Setelah itu, metode dengan skema terbaik akan diimplementasikan pada sebuah aplikasi *web* sederhana berbasis *framework Streamlit*. Aplikasi ini akan menerima rekaman suara vokal manusia (akapela) sebagai masukan dan selanjutnya akan diprediksi menggunakan model terbaik untuk menghasilkan *output* berupa suara baru yang menyerupai iringan sebuah instrumen musik. Tujuan dari penelitian ini adalah untuk mengembangkan sebuah sistem otomatis yang dapat memproduksi musik instrumental secara otomatis dari rekaman suara manusia, dan untuk mengevaluasi metode yang paling efektif dalam menghasilkan *output* musik instrumental yang berkualitas.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, penulis tertarik untuk melakukan penelitian tentang pembuatan sebuah model yang dapat memprediksi suatu

masukan suara vokal manusia yang bernada untuk menghasilkan *output* berupa sebuah suara iringan instrumen musik menggunakan teknik *deep learning* seperti *autoencoder* dan RNN dengan memanfaatkan proses *preprocessing* data sisir (*combing*) dan DCT.

### 1.3 Pertanyaan Penelitian

Berdasarkan rumusan masalah di atas, maka pertanyaan yang muncul pada penelitian ini adalah :

1. Bagaimana implementasi teknik *autoencoder* dan RNN pada studi kasus pembuat instrumen musik otomatis berdasarkan suara vokal manusia?
2. Bagaimana membuat arsitektur model *autoencoder* yang baik untuk mendapatkan *latent* dengan ukuran optimal dan ketika diprediksi menggunakan model RNN, *output* yang dihasilkan masih dapat didengar?
3. Berapa *layer* yang dibutuhkan pada rekonstruksi model arsitektur model *autoencoder* dan model RNN untuk hasil evaluasi terbaik?
4. Apakah teknik sisir (*combing*) dan DCT dapat menjadi representasi yang baik untuk *preprocessing* data suara agar dapat diproses pada spesifikasi komputer rendah?
5. Bagaimana hasil *test* dengan *output* dari metode skema *Autoencoder* – RNN, *Autoencoder* - Sisir (*Combing*) - RNN, dan DCT - RNN?

### 1.4 Tujuan Penelitian

Berdasarkan pertanyaan penelitian di atas, maka tujuan dilakukannya penelitian ini adalah :

1. Untuk mengimplementasikan teknik *autoencoder* dan RNN pada studi kasus pembuat instrumen musik otomatis berdasarkan suara vokal manusia.

2. Untuk membuat arsitektur model *autoencoder* yang baik agar dapat menghasilkan *latent* dengan ukuran optimal dan ketika diprediksi menggunakan model RNN *output* yang dihasilkan masih dapat didengar.
3. Untuk menentukan jumlah *layer* yang optimal pada rekonstruksi model arsitektur model *autoencoder* dan model RNN agar menghasilkan evaluasi terbaik.
4. Untuk mengevaluasi teknik sisir (*combing*) dan DCT dapat menjadi representasi yang baik untuk *preprocessing* data suara agar dapat diproses pada spesifikasi komputer rendah.
5. Untuk membandingkan hasil *test* dengan *output* dari metode skema *Autoencoder – RNN*, *Autoencoder - Sisir (Combing) - RNN*, dan *DCT - RNN* sehingga dapat mengetahui teknik mana yang paling efektif dalam menghasilkan suara instrumen yang bagus.

## 1.5 Batasan Masalah

Adapun batasan masalah yang terdapat pada penelitian ini adalah sebagai berikut:

1. *Dataset* yang digunakan pada penelitian ini berupa suara instrumen musik gitar dan suara vokal manusia pada sebuah lagu *cover* akustik.
2. Penelitian ini menghasilkan *output* berupa instrumen musik gitar yang *tergenerate*.
3. *Dataset* lagu yang *ditrain* dan *ditest* adalah berbahasa Inggris.
4. *Dataset* instrumen musik gitar dan suara vokal (lagu) yang akan digunakan berada pada periode tahun 1990-2022.
5. *Genre* musik yang dipilih adalah Pop remaja-dewasa.
6. Model dengan evaluasi terbaik akan diimplementasikan pada aplikasi berbasis *web* menggunakan *framework streamlit*.

## **1.6 Manfaat Penelitian**

Penelitian ini diharapkan dapat memberikan manfaat yang signifikan bagi mereka yang kurang berpengalaman dalam bidang musik. Dengan adanya model yang dapat memprediksi irama instrumen musik yang harmonis dari suara vokal manusia, diharapkan akan mempermudah mereka dalam mengekspresikan ide kreatif mereka dalam menciptakan instrumen musik yang unik. Selain itu, penelitian ini juga berpotensi untuk mendorong pengembangan lebih lanjut dalam bidang penggabungan antara musik dan teknologi. Potensi tersebut sangat besar di masa depan karena semakin banyak inovasi teknologi yang dapat dimanfaatkan untuk menghasilkan karya musik yang lebih bervariasi dan berkualitas tinggi. Dengan demikian, hasil penelitian ini diharapkan dapat memberikan kontribusi positif bagi perkembangan industri musik di masa depan.