

Pengembangan Model Deep Learning Long Short-Term Memory untuk Generasi Musik Berbasis Data MIDI

Muhammad Djamaruddin¹, Imam Fathurrahman^{2*}, M. Nurul Wathani³

^{1,2,3}Program Studi Informatika, Universitas Hamzanwadi

^{*}i.fathurrahman@hamzanwadi.ac.id

Abstrak

Penelitian ini mengembangkan model generasi musik otomatis berbasis Long Short-Term Memory (LSTM) dengan memanfaatkan data MIDI sebagai representasi simbolik dari rangkaian musik piano klasik. Pendekatan yang digunakan bersifat eksperimen komputasional, dengan alur kerja yang mencakup tahapan ekstraksi dan konversi file MIDI menggunakan music21, pembentukan token nada dan akor, penyusunan sekuens input-output, perancangan arsitektur tiga lapis LSTM, serta proses pembangkitan musik secara autoregresif. Model dilatih selama 100 epoch dengan batch size sebesar 64 dan dievaluasi menggunakan metrik loss, accuracy, top-3 accuracy, serta perplexity untuk menilai kemampuan prediktif terhadap data yang tidak dilatih. Hasil pengujian menunjukkan penurunan validation loss yang konsisten, dengan nilai akhir sekitar 2,93 dan validation accuracy sebesar 0,33, sedangkan top-3 accuracy mencapai 0,53, yang mengindikasikan bahwa lebih dari separuh prediksi benar berada dalam tiga kemungkinan teratas. Nilai perplexity yang berada di kisaran 18 menunjukkan bahwa model memiliki kemampuan prediksi sekuens yang cukup memadai untuk konteks data musik simbolik. Secara kualitatif, model mampu menghasilkan melodi sederhana dengan pola yang masih koheren terhadap distribusi nada pada dataset, meskipun beberapa bagian komposisi masih menunjukkan repetisi dan keterbatasan variasi. Kontribusi penting dari penelitian ini adalah tersusunnya dokumentasi metodologis yang sistematis terkait pipeline generasi musik berbasis LSTM, sehingga dapat menjadi referensi praktis bagi pengembangan dan penelitian lanjutan di bidang pembangkitan musik berbasis deep learning.

Kata kunci : Deep Learning, Generasi Musik, LSTM, MIDI, Model Sekuens

Abstract

This study develops an automatic music generation model based on Long Short-Term Memory (LSTM) by utilizing MIDI data as a symbolic representation of classical piano music sequences. The approach is computational and experimental, with a workflow that includes extracting and converting MIDI files using music21, constructing note and chord tokens, forming input-output sequences, designing a three-layer LSTM architecture, and generating music in an autoregressive manner. The model is trained for 100 epochs with a batch size of 64 and evaluated using loss, accuracy, top-3 accuracy, and perplexity metrics to assess its predictive capability on unseen data. The experimental results show a consistent decrease in validation loss, with a final value of approximately 2.93 and a validation accuracy of 0.33, while the top-3 accuracy reaches 0.53, indicating that more than half of the correct predictions fall within the top three candidates. A perplexity value around 18 suggests that the model has a reasonably adequate sequence prediction ability for symbolic music data. Qualitatively, the model is able to generate simple melodies whose patterns remain coherent with the note distribution in the dataset, although some parts of the compositions still exhibit repetition and limited variation. An important contribution of this study is the provision of a systematic methodological documentation of the LSTM-based music generation pipeline, which can serve as a practical reference for future development and research in deep learning-based music generation.

Keywords : Deep Learning, Music Generation, LSTM, MIDI, Sequence Model.

1. Pendahuluan

Musik memiliki posisi penting sebagai media ekspresi yang bersifat universal dan mewarnai banyak aspek kehidupan manusia. Seiring kemajuan teknologi, muncul peluang untuk mengeksplorasi cara baru dalam menciptakan musik, salah satunya melalui sistem generatif berbasis komputer. Di tengah pesatnya perkembangan bidang ini, tantangan utama tetap pada upaya menjaga kesinambungan struktur melodi dan harmoni, sehingga hasil komposisi tetap terdengar alami dan koheren [1], [2].

Di sisi lain, pemanfaatan kecerdasan buatan, khususnya teknik machine learning dan deep learning berbasis jaringan saraf tiruan (JST), telah banyak dieksplorasi pada berbagai domain lain. Contohnya, penelitian Fathurrahman dkk. pada pengenalan citra dan klasifikasi citra berbasis JST-Backpropagation maupun CNN menunjukkan bahwa model jaringan saraf mampu mempelajari pola kompleks secara otomatis, sehingga secara metodologis berpotensi diadaptasi ke domain data sekuensial seperti musik. [3], [4], [5], [6].

Metode-metode tradisional seperti *rule-based* dan *Markov Chains* memang telah lama digunakan, namun belum mampu memenuhi kebutuhan akan variasi serta koherensi dalam generasi komposisi yang kompleks. Oleh sebab itu, berbagai penelitian mulai memanfaatkan pendekatan kecerdasan buatan, khususnya teknik *machine*

learning berbasis *deep learning* dengan JST berlapis banyak yang dirancang untuk memproses data berurutan, seperti *Recurrent Neural Network* (RNN) dan *Long Short-Term Memory* (LSTM). Model ini dikenal efektif dalam mempelajari pola-pola musical berdurasi panjang karena kemampuannya memahami konteks yang luas dan melakukan prediksi berurutan yang lebih stabil [1], [7], [8], [9], [10].

Namun demikian, penelitian - penelitian sebelumnya masih terbatas pada penerapan LSTM secara umum tanpa kajian yang terstruktur mengenai bagaimana representasi sekuens berbasis data MIDI memengaruhi kemampuan model dalam menjaga koherensi melodi dan harmoni pada proses generasi musik otomatis. Pemanfaatan file MIDI sebagai media representasi digital musik menjadi pilihan yang tepat karena menyimpan informasi penting berupa nada dan harmoni. Proses pengolahan data dalam penelitian ini meliputi ekstraksi nada dan chord, pembuatan daftar token unik, hingga pembentukan urutan data untuk pelatihan model. Selanjutnya, hasil pelatihan model digunakan untuk membangkitkan urutan nada dan harmoni baru secara bertahap, dengan mekanisme pengambilan keputusan yang mempertimbangkan unsur acak agar hasilnya variatif namun tetap berada dalam koridor pola data aslinya [1], [7], [11].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk membangun dan mengevaluasi model generasi musik otomatis berbasis *Long Short-Term Memory* (LSTM) dengan memanfaatkan dataset MIDI sebagai representasi sekuens musik. Melalui pendekatan eksperimen komputasional, penelitian ini diharapkan mampu menghasilkan model yang dapat mempelajari pola musical secara efektif dan menghasilkan komposisi baru yang tetap menjaga koherensi nada dan struktur melodi. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi pada pengembangan teknologi generatif berbasis deep learning, tetapi juga menyediakan dokumentasi terstruktur mengenai proses implementasi, pelatihan, dan evaluasi model generasi musik otomatis..

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Beberapa penelitian terdahulu telah memberikan kontribusi penting dalam pengembangan sistem generasi musik berbasis deep learning, sbb:

- Wanjari *et al.* mengembangkan sistem generasi musik otomatis berbasis LSTM menggunakan dataset Piano MIDI sebagai data latih. Penelitian ini melakukan ekstraksi nada dan chord dari file MIDI, membentuk urutan token bernada tetap, lalu melatih model LSTM untuk memprediksi token berikutnya secara autoregresif sehingga mampu menghasilkan rangkaian nada

baru dengan struktur yang menyerupai data latihan. Hasil eksperimen menunjukkan bahwa model LSTM mencapai akurasi tinggi (hingga sekitar 97,23% pada dataset Piano MIDI setelah sejumlah epoch pelatihan), yang mengindikasikan kemampuan model dalam mempertahankan koherensi melodi dan harmoni selama proses generasi [7].

- Han *et al.* mengusulkan model generasi musik yang digerakkan oleh emosi (*emotion-driven*) berbasis deep learning, di mana jaringan LSTM digunakan untuk memetakan representasi emosi ke sekuens nada musik. Model dilatih dengan data musik beranotasi emosi, kemudian menghasilkan komposisi baru dengan mengendalikan label emosi sebagai kondisi input. Penelitian ini menegaskan bahwa LSTM efektif dalam mempelajari hubungan jangka panjang pada sekuens musik, sekaligus menunjukkan bahwa mekanisme generatif *autoregresif* dapat diarahkan untuk menghasilkan karakter emosional tertentu pada musik [12].

- Lysal *et al.* Dalam penelitiannya berfokus pada penerapan jaringan syaraf berulang, khususnya LSTM, untuk menyusun rangkaian nada dan akor secara otomatis sebagai bentuk komposisi musik generatif. Tujuan utamanya adalah menunjukkan bagaimana model LSTM dapat mempelajari pola sekuensial dalam data musik serta mendemonstrasikan peran kecerdasan buatan dalam mendukung kreativitas di bidang musik [1].

- Huang *et al.* mengembangkan model generasi musik berbasis Convolution-LSTM yang menggabungkan CNN dan LSTM untuk menangani musik polyphonic. CNN digunakan untuk mengekstraksi fitur spasial dari representasi piano-roll, sedangkan LSTM menangkap dependensi temporal antar frame, sehingga model dapat mempelajari hubungan simultan (vertikal) dan sekuensial (horizontal) pada musik. Hasil penelitian menunjukkan bahwa model CNN-LSTM menghasilkan pola harmoni dan undulasi melodi yang lebih jelas serta struktur musik yang lebih koheren dibanding LSTM murni [10].

- Dalam karya lain, Huang *et al.* memperkenalkan Musik *Transformer* yang menggunakan mekanisme *self-attention* untuk generasi musik dengan struktur jangka panjang yang kuat. Berbeda dari LSTM yang mengandalkan memori berurutan, Musik *Transformer* memanfaatkan *self-attention* untuk mengakses konteks global pada sekuens musik, sehingga mampu menghasilkan komposisi berdurasi panjang dengan motif yang konsisten dan pengembangan tema yang lebih terstruktur. Penelitian ini menjadi salah satu rujukan utama model *Transformer* sebagai *state-of-the-art* dalam generasi musik otomatis [13].

2.2. Landasan Teori

1 Musik Digital

Musik digital merupakan representasi musik dalam format digital, memungkinkan analisis, komposisi, dan generasi musik secara otomatis melalui komputer dan algoritma pembelajaran mesin. Format digital telah membawa perubahan besar dalam industri musik Indonesia dengan migrasi dari media fisik ke distribusi digital dan streaming, meningkatkan akses dan efisiensi produksi serta pemasaran musik secara nasional^[14].

2 MIDI (*Musical Instrument Digital Interface*) adalah protokol dan format file yang merepresentasikan musik sebagai rangkaian event digital, bukan suara gelombang analog. Dalam MIDI, informasi nada, waktu, durasi, velocity, dan channel instrumen dikodekan secara diskrit, sehingga sangat sesuai digunakan dalam proses pembangkitan musik otomatis berbasis komputer, baik untuk keperluan analisis, komposisi, maupun pembelajaran musik digital [15], [16]

3 LSTM

LSTM (*Long Short - Term Memory*) merupakan jenis *Recurrent Neural Network* (RNN) yang memiliki kemampuan memori jangka panjang lewat mekanisme *input*, *forget*, dan *output gate*. Keunggulan utama LSTM adalah kemampuannya mengatasi masalah *vanishing gradient* pada RNN tradisional dan secara akurat belajar pola data

sekuensial serta *dependensi temporal*. Dalam konteks musik, LSTM dapat digunakan untuk mempelajari dan menghasilkan rangkaian nada atau data musical yang menyerupai struktur aslinya [15], [17], [18].

4 Generasi Musik Otomatis

Generasi musik otomatis adalah proses komputer secara mandiri untuk menghasilkan komposisi baru menggunakan algoritma kecerdasan buatan (AI) dan data musik, seperti MIDI. Sistem ini memungkinkan penciptaan musik baru dengan pola dan karakteristik yang menyerupai data sumber, tanpa interaksi langsung dari manusia, serta menjadi kunci perkembangan sistem rekomendasi, klasifikasi, dan personalisasi musik digital di Indonesia [19], [20].

5 Deep learning

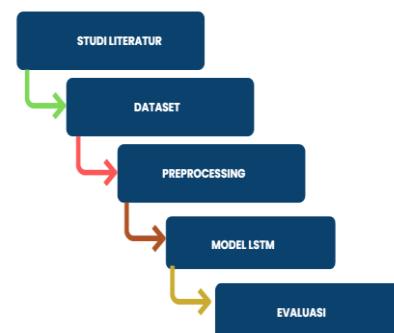
Deep learning merupakan cabang dari *machine learning* yang memanfaatkan jaringan saraf tiruan berlapis banyak untuk merepresentasikan data pada berbagai tingkat abstraksi sehingga mampu membangun model komputasi yang efektif untuk tugas visual, audio, teks, dan data sekuensial lainnya. Dalam hierarki kecerdasan buatan, *deep learning* dipandang sebagai subset dari *machine learning* yang berada di bawah payung besar *artificial intelligence* (AI), dan telah menjadi pendekatan utama dalam berbagai aplikasi modern melalui algoritma seperti *convolutional neural networks* dan *generative adversarial networks* [21].

3. Metode Penelitian

Penelitian ini menggunakan pendekatan eksperimen Komputasional untuk membangun dan mengevaluasi model generasi musik otomatis berbasis LSTM. Pendekatan ini dipilih karena LSTM mampu mempelajari pola data sekuensial secara efektif, khususnya pada konteks musik yang memiliki dependensi temporal panjang [1][2][10].

3.1. Tahapan Penelitian

Tahapan-tahapan dalam penelitian ini dapat digambarkan secara visual pada Gambar 1.



Gambar 1. Tahapan Penelitian

3.2. Studi Literatur

Penelitian diawali dengan studi literatur terkait konsep musik digital, representasi data musik dalam format MIDI, serta perkembangan metode pembelajaran mesin untuk pemodelan sekuens, khususnya arsitektur *Recurrent Neural Network* (RNN) dan *Long Short-Term Memory* (LSTM). Selain itu, telaah terhadap penelitian sebelumnya dilakukan untuk mengidentifikasi perkembangan generasi musik otomatis, keunggulan dan keterbatasan metode LSTM, serta gap riset terkait

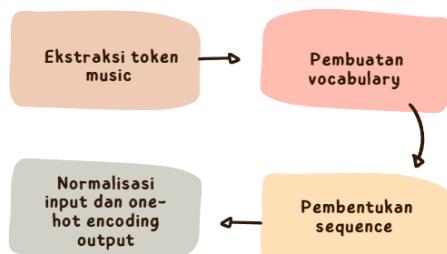
kebutuhan dokumentasi proses implementasi generatif berbasis MIDI. Studi literatur ini menjadi dasar dalam merumuskan desain model, teknik preprocessing, dan metode evaluasi yang digunakan dalam penelitian.

3.3. Pengumpulan dan Persiapan Dataset

Penelitian ini menggunakan kumpulan file MIDI piano klasik yang diperoleh dari sumber publik <https://www.kaggle.com/datasets/soumikrakshit/classical-music-midi>. Terdapat 295 file MIDI dan seluruh file diekstraksi menggunakan library *music21* untuk memperoleh elemen-elemen musik yang diperlukan. Dataset ini terdiri atas berbagai karya piano klasik dengan struktur melodi yang jelas, sehingga sangat sesuai untuk pelatihan model berbasis urutan (*sequence*).

Setiap file MIDI diparsing menggunakan fungsi *instrument.partitionByInstrument* untuk mengambil partitur piano utama. Elemen yang diambil berupa nada tunggal (*pitch*) dan *chord*. Hasil ekstraksi *pitch* dan *chord* tersebut selanjutnya digunakan sebagai token musik dalam proses pelatihan model.

3.4. Preprocessing



Gambar 2. Tahapan preprocessing

Tahapan preprocessing dilakukan untuk mengubah data musik menjadi format numerik yang dapat diproses oleh model LSTM. Terdiri sbb:

1. Ekstraksi token musik yaitu Setiap *not* atau *chord* disimpan sebagai token string kemudian seluruh token dikumpulkan dalam daftar notes.
2. Pembuatan *vocabulary* yaitu Token unik dari seluruh dataset diurutkan dan dikonversi menjadi indeks numerik menggunakan *mapping*
3. Pembentukan *sequence* untuk pelatihan yaitu Model LSTM memerlukan input berupa *sequence* tetap menggunakan *sliding windows*
4. Normalisasi input dan *one-hot encoding output*, menghasilkan *training-ready sequence* yang digunakan pada pemodelan LSTM

3.5. Arsitektur Model LSTM

Model generatif dibangun menggunakan kerangka kerja Keras secara bertingkat (*stacked LSTM*). Arsitektur yang digunakan adalah LSTM 3 layer masing masing 256 unit dan dropout 0.3. Menggunakan *softmax* untuk menghasilkan distribusi probabilitas untuk tiap token musik.

3.6. Pelatihan Model

Model dilatih menggunakan parameter sebagai berikut: *Epoch*: 100, *Batch size*: 64, *Optimizer*: Adam, *Loss function*: categorical crossentropy, *Validation* : 20%, *Testing* : 80%

Selama pelatihan, model mempelajari hubungan antar token sehingga dapat memprediksi token berikutnya secara autoregresif. Proses pelatihan

dilakukan pada seluruh pasangan input-output yang dihasilkan dari preprocessing.

3.7. Proses Generasi Musik

Setelah model selesai dilatih, dilakukan proses generasi musik secara *autoregresif*. Proses ini menghasilkan komposisi musik baru yang mengikuti karakter pola melodi dari dataset pelatihan.

3.8. Evaluasi Model

Setelah proses pelatihan selesai, model dievaluasi menggunakan beberapa metrik, yaitu:

1. Validation Loss

Validation loss merupakan metrik utama yang digunakan untuk mengevaluasi tingkat kesalahan prediksi model pada *validation set*. Pengukuran ini menggunakan fungsi *categorical cross-entropy*, yang menghitung perbedaan antara distribusi probabilitas prediksi model dan distribusi target sebenarnya. Nilai loss yang rendah menunjukkan bahwa model mampu mempelajari pola data secara efektif tanpa mengalami *overfitting*. Tren penurunan *validation loss* selama proses pelatihan mencerminkan peningkatan kemampuan generalisasi model terhadap data baru yang tidak terlihat sebelumnya.

2. Validation Accuracy

Validation accuracy mengukur persentase prediksi model yang sesuai dengan label target pada data validasi. Metrik ini memberikan gambaran mengenai kemampuan model dalam memilih token yang benar di antara seluruh

kemungkinan token dalam *vocabulary*. Akurasi yang tinggi menunjukkan bahwa representasi temporal dan pola musical yang dipelajari oleh model telah selaras dengan data pelatihan. Sebaliknya, bila akurasi tinggi pada data pelatihan tetapi rendah pada data validasi, hal tersebut mengindikasikan *overfitting*.

3. Top-3 Accuracy

Top-3 accuracy mengevaluasi apakah token target berada dalam tiga prediksi dengan probabilitas tertinggi yang dihasilkan model. Metrik ini sangat relevan untuk sistem generatif musik, karena dalam banyak kasus model tidak hanya diharapkan untuk membuat satu prediksi yang benar secara deterministik, tetapi juga menangkap kemungkinan alternatif yang masih musical dan valid. Dengan demikian, top-3 accuracy memberikan gambaran yang lebih fleksibel mengenai pemahaman model terhadap distribusi musik, terutama pada situasi di mana beberapa token dapat diterima sebagai kelanjutan yang musical.

4. Perplexity

Perplexity adalah metrik fundamental dalam evaluasi model berbasis urutan (*sequence models*), termasuk model generatif musik. Perplexity mengukur tingkat ketidakpastian model ketika memprediksi token berikutnya dalam sebuah urutan. Nilai ini dihitung menggunakan rumus:

$$\text{Perplexity} = e^{-\frac{1}{N} \sum_{i=1}^N \log P(x_i)}$$

Keterangan:

- N = jumlah token dalam urutan (sequence)
- $P(x_i)$ = probabilitas prediksi model terhadap token ke- i
- \log = logaritma natural (basis e)

Perplexity yang rendah menunjukkan bahwa model memiliki keyakinan yang tinggi terhadap prediksi yang dihasilkan, serta mampu mempelajari distribusi probabilitas token secara stabil. Sebaliknya, *perplexity* yang tinggi mengindikasikan bahwa model mengalami kesulitan dalam memprediksi token berikutnya, baik akibat data yang kompleks maupun representasi yang belum optimal. Dalam penelitian ini, *perplexity* digunakan untuk menilai kualitas probabilistik model secara menyeluruh, tidak hanya sebatas ketepatan prediksi tunggal.

3.8 Lokasi Penelitian

1. Penelitian ini dilakukan di Laboratorium Informatika, Fakultas Teknik, Universitas Hamzanwadi pada periode Agustus hingga Oktober 2025.

4. Hasil dan Pembahasan

4.1. Hasil Penelitian

1. Persiapan Dataset

Dataset yang digunakan dalam penelitian ini diperoleh dari platform Kaggle ([kaggle.com](https://www.kaggle.com)) dalam bentuk berkas terkompresi dengan ekstensi .zip. Dataset tersebut berisi kumpulan berkas musik digital berformat MIDI (.mid) yang

umumnya digunakan dalam penelitian terkait musik karena dapat merepresentasikan struktur musical secara simbolik, mencakup *pitch*, durasi, *chord*, dan elemen komposisi lainnya.

Name	Type	Compressed size
alb_esp1.mid	MIDI Sequence	4 KB
alb_esp2.mid	MIDI Sequence	4 KB
alb_esp3.mid	MIDI Sequence	5 KB
alb_esp4.mid	MIDI Sequence	6 KB
alb_esp5.mid	MIDI Sequence	4 KB
alb_esp6.mid	MIDI Sequence	4 KB
alb_se1.mid	MIDI Sequence	6 KB
alb_se2.mid	MIDI Sequence	7 KB
alb_se3.mid	MIDI Sequence	10 KB

Gambar 3. Dataset MIDI

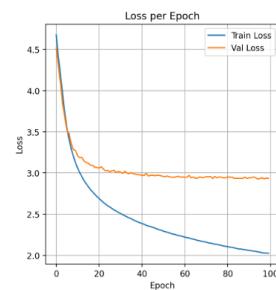
Setelah proses ekstraksi dilakukan, seluruh berkas MIDI kemudian dikumpulkan ke dalam satu direktori kerja agar mudah diakses selama proses pemrosesan data. Proses ekstraksi dan pemutuan awal dataset bertujuan memastikan setiap berkas MIDI dapat dibaca oleh *library music21* yang digunakan dalam tahap preprocessing.

2. Evaluasi Model

Untuk mengevaluasi kinerja model secara menyeluruh, dilakukan kombinasi evaluasi kuantitatif menggunakan metrik statistik

Pembahasan berikut disajikan berdasarkan empat indikator utama yaitu:

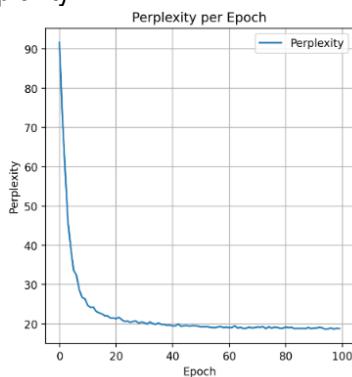
- Validation Loss



Gambar 4. Validation Loss

Gambar 4 menampilkan grafik *Loss per Epoch* menunjukkan bahwa nilai *training loss* mengalami penurunan yang konsisten dari awal hingga akhir pelatihan. Pada awal pelatihan, *training loss* berada di atas 4.5 dan menurun hingga mencapai sekitar 2.0 pada epoch ke-100. Penurunan ini menunjukkan bahwa model berhasil mempelajari pola sekuens nada secara progresif. Sementara itu, *validation loss* juga mengalami penurunan, namun tidak sedrastis *training loss*. Nilai *validation loss* cenderung stabil di sekitar 2.9 setelah memasuki pertengahan pelatihan. Perbedaan antara *training loss* dan *validation loss* menunjukkan adanya kecenderungan *overfitting ringan*, namun masih dalam batas yang dapat diterima untuk model generatif musik.

- **Perplexity**



Gambar 5. Perplexity

Perplexity merupakan metrik penting dalam pemodelan sekuens karena mengukur seberapa baik model memprediksi token berikutnya. Pada gambar 5 Grafik *Perplexity per Epoch* menunjukkan tren penurunan yang jelas sepanjang pelatihan. Pada awal pelatihan

perplexity berada di atas 90, kemudian mengalami penurunan tajam hingga mencapai nilai sekitar 18 pada akhir pelatihan.

Nilai *perplexity* yang rendah menandakan bahwa model semakin yakin dalam menghasilkan prediksi not berikutnya dan mampu menangkap pola probabilistik yang muncul dalam data musik secara efektif.

- **Top-3 Accuracy**

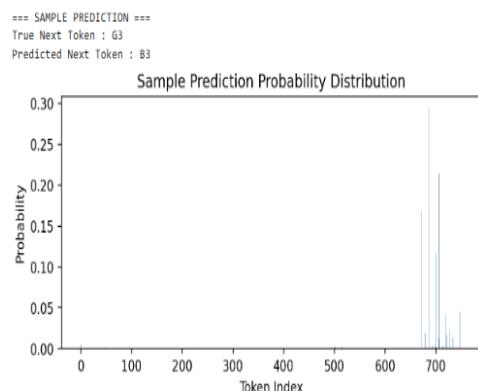
<i>Validation Loss</i>	:	2.9341
<i>Validation Accuracy</i>	:	0.3397
<i>Top-3 Accuracy</i>	:	0.5385
<i>Perplexity</i>	:	18.80

Gambar 6. Top-3 Accuracy

Terlihat pada gambar 6 *Top-3 accuracy* pada model mencapai nilai 0.5385, yang berarti bahwa sekitar 53.85% prediksi yang benar berada di dalam tiga kemungkinan teratas yang dihasilkan model. Metrik ini sangat relevan untuk pemodelan musik, karena satu nada berikutnya tidak selalu memiliki satu jawaban deterministik, melainkan beberapa kemungkinan logis berdasarkan konteks harmonik dan ritmik. Oleh karena itu, evaluasi berbasis *Top-3* lebih mencerminkan kemampuan model dalam memahami struktur musik secara menyeluruh dibandingkan hanya mengukur prediksi tunggal yang tepat. Nilai *Top-3 accuracy* tersebut menunjukkan bahwa model mampu menangkap pola musical dengan cukup baik, karena lebih dari separuh prediksi yang benar muncul sebagai salah satu dari tiga

kandidat teratas dalam distribusi probabilitas output.

- **Analisis Distribusi Probabilitas Prediksi**



Gambar 7. Probabilitas Prediksi

Gambar 7 *Sample Prediction Probability Distribution* menunjukkan bahwa model menghasilkan distribusi probabilitas yang sangat terfokus pada beberapa token tertentu. Titik-titik probabilitas yang tinggi menandakan bahwa model telah mempelajari transisi nada yang dominan dalam dataset.

== SAMPLE PREDICTION ==
True Next Token : G3
Predicted Next Token : B3

Gambar 8. Sample prediction

Dapat dilihat pada gambar 8 prediksi terdapat:

- ✓ *True next token: G3*
- ✓ *Predicted token: B3*

Perbedaan ini menunjukkan bahwa model tidak selalu menghasilkan prediksi yang tepat, namun tetap memberikan distribusi probabilitas yang masuk akal secara musical. Hal ini sangat umum dalam model generatif musik, mengingat sifat

data yang memiliki banyak variasi harmonis dan ritmis.

4.2. Pembahasan

Secara umum, hasil pelatihan menunjukkan bahwa : (1) Model LSTM berhasil mempelajari pola musik berdasarkan penurunan *loss* dan *perplexity*. (2) Akurasi validasi yang stabil menandakan kemampuan generalisasi yang cukup baik meskipun masih terdapat jarak dengan *training accuracy*. (3) *Perplexity* yang rendah menegaskan bahwa model efektif dalam memprediksi token musik berikutnya. (4) Distribusi probabilitas prediksi menunjukkan tingkat keyakinan model yang cukup tinggi terhadap nada tertentu, meskipun tidak selalu menghasilkan prediksi yang identik dengan data asli. Secara keseluruhan, performa model dapat dikategorikan cukup baik untuk sistem generatif musik berbasis LSTM, dan hasil ini dapat menjadi dasar untuk pengembangan model lanjutan seperti penambahan lapisan, penggunaan mekanisme *attention*, dan model dapat dikembangkan menggunakan Transformer atau dataset musik yang lebih besar.

5. Kesimpulan

Berdasarkan hasil eksperimen dan evaluasi model, dapat disimpulkan bahwa arsitektur LSTM yang digunakan mampu mempelajari pola sekuens musik dengan cukup baik. Hal ini terlihat dari nilai validation loss sebesar 2,9341 dan

validation accuracy sebesar 0,3397, yang mengindikasikan kemampuan generalisasi model yang moderat terhadap data yang tidak dilatih. Nilai top-3 accuracy sebesar 0,5385 menunjukkan bahwa lebih dari separuh prediksi yang benar berada di antara tiga kemungkinan teratas, sehingga model mampu menangkap variasi transisi nada secara lebih fleksibel dan tidak sekadar menghafal pola tunggal.

Kontribusi utama penelitian ini adalah tersusunnya pipeline lengkap generasi musik otomatis berbasis LSTM menggunakan data MIDI, mulai dari tahap pra-pemrosesan (ekstraksi nada dan chord, pembentukan token dan sekuens), perancangan serta pelatihan model, hingga proses generasi dan rekonstruksi kembali menjadi file MIDI yang dapat didengarkan. Dokumentasi terstruktur ini memberikan acuan praktis bagi peneliti dan pengembang lain yang ingin mereplikasi atau mengembangkan lebih lanjut sistem generasi musik berbasis LSTM, misalnya dengan memperluas arsitektur, menambah mekanisme *conditioning*, atau mengintegrasikan metode evaluasi subjektif berbasis pendapat pendengar untuk menilai kualitas musical secara lebih menyeluruh..

6. Daftar Pustaka

- [1] A. S. J. Lysal, M. P. Jothilakshmi, P. Muralidharan, and S. S. Rathipriya, "Generation of music using LSTM," 2025, p. 020022. doi: 10.1063/5.0263106.
- [2] M. Han, S. Soradi-Zeid, T. Anwlnkom, and Y. Yang, "Firefly algorithm-based LSTM model for Guzheng tunes switching with big data analysis," *Heliyon*, vol. 10, no. 12, p. e32092, Jun. 2024, doi: 10.1016/j.heliyon.2024.e32092.
- [3] I. Fathurrahman and I. Gunawan, "Pengenalan Citra Logo Kendaraan Menggunakan Metode Gray Level Co-Occurrence Matrix (Glcm) dan Jst-Backpropagation," *Infotek : Jurnal Informatika dan Teknologi*, vol. 1, no. 1, pp. 47–55, Jan. 2018, doi: 10.29408/jit.v1i1.894.
- [4] I. Fathurrahman, A. M. Nur, and F. Farhurrahman, "Identifikasi Kematangan Buah Mentimun Berbasis Citra Digital Menggunakan Jaringan Syaraf Tiruan Backpropagation," *Infotek: Jurnal Informatika dan Teknologi*, vol. 2, no. 1, pp. 27–33, Jan. 2019, doi: 10.29408/jit.v2i1.976.
- [5] I. Fathurrahman, M. Djamiluddin, Z. Amri, and M. N. Wathan, "Klasifikasi Motif Batik Nusantara Menggunakan Vision Transformer (ViT) Berbasis Deep Learning," *Infotek: Jurnal Informatika dan Teknologi*, vol. 8, no. 2, pp. 511–522, Jul. 2025, doi: 10.29408/jit.v8i2.31108.
- [6] Imam Fathurrahman, Mahpuz, Muhammad Djamiluddin, Lalu Kerta Wijaya, and Ida Wahidah, "Pengembangan Model Convolutional Neural Network (CNN) untuk Klasifikasi Penyakit Kulit Berbasis Citra Digital," *Infotek: Jurnal Informatika dan Teknologi*, vol. 8, no. 1, pp. 298–308, Jan. 2025, doi: 10.29408/jit.v8i1.28655.
- [7] S. Wanjari, P. Tupe, A. Nawale, and P. Papade, "Music Generation Using LSTM Model," *TIJER - International Research Journal*, vol. 10, no. 12, Dec. 2023, [Online]. Available: www.tijer.org
- [8] S. S. Patil *et al.*, "Music Generation Using RNN-LSTM with GRU," in 2023 *International Conference on Integration of Computational Intelligent System (ICICIS)*,

- IEEE, Nov. 2023, pp. 1–5. doi: 10.1109/ICICIS56802.2023.10430293.
- [9] A. Ycart and E. Benetos, “A Study on Lstm Networks For Polyphonic Music Sequence Modelling,” China, Aug. 2017. [Online]. Available: <http://www.eecs.qmul.ac.uk/>
- [10] Y. Huang, X. Huang, and Q. Cai, “Music Generation Based on Convolution-LSTM,” *Computer and Information Science*, vol. 11, no. 3, p. 50, Jun. 2018, doi: 10.5539/cis.v11n3p50.
- [11] T. Do Quang and T. Hoang, “An efficient method to build music generative model by controlling both general and local note characteristics,” *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 9, p. 101761, Oct. 2023, doi: 10.1016/j.jksuci.2023.101761.
- [12] Z. Ning, X. Han, and J. Pan, “Semi-supervised emotion-driven music generation model based on category-dispersed Gaussian Mixture Variational Autoencoders,” *PLoS One*, vol. 19, no. 12, p. e0311541, Dec. 2024, doi: 10.1371/journal.pone.0311541.
- [13] C.-Z. A. Huang *et al.*, “Music Transformer: Generating Music with Long-Term Structure,” in *International Conference on Learning Representations*, 2018. [Online]. Available: <https://api.semanticscholar.org/CorpusID:54477714>
- [14] I. Ruddin, H. Santoso, and R. E. Indrajit, “Digitalisasi Musik Industri: Bagaimana Teknologi Informasi Mempengaruhi Industri Musik di Indonesia,” *Jurnal Pendidikan Sains dan Komputer*, vol. 2, no. 01, pp. 124–136, Feb. 2022, doi: 10.47709/jpsk.v2i01.1395.
- [15] T. S. Fitriani, A. Saepudin, and J. Karawitan, “Midi Sebagai Inovasi Dan Alternatif Musik Iringan Tari Di Masa Pandemi,” vol. 5, no. 1, 2022, doi: 10.26887/mapj.
- [16] Y. Samosir, I. K. G. Suhartana, and I. G. N. A. C. Putra, “Konversi Suara ke MIDI Menggunakan Short Time Fourier Transform Sebagai Virtual Midi Controller Pada Digital Audio Workstation,” *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, vol. 12, no. 2, p. 451, Aug. 2023, doi: 10.24843/JLK.2023.v12.i02.p24.
- [17] J. Cahyani, S. Mujahidin, and T. P. Fiqar, “Implementasi Metode Long Short Term Memory (LSTM) untuk Memprediksi Harga Bahan Pokok Nasional,” *Jurnal Sistem dan Teknologi Informasi (JustIN)*, vol. 11, no. 2, p. 346, Jul. 2023, doi: 10.26418/justin.v11i2.57395.
- [18] M. A. Saputra and T. Sugihartono, “Evaluasi Kinerja Model LSTM Untuk Prediksi Risiko Penyakit Jantung Menggunakan Dataset,” *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 7, pp. 1823–1833, Jul. 2025, doi: 10.52436/1.jpti.821.
- [19] D. ROSMALA and M. N. FADHILAH, “Audio Conversion for Music Genre Classification Using Short-Time Fourier Transform and Inception V3,” *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, vol. 13, no. 1, p. 84, Jan. 2025, doi: 10.26760/elkomika.v13i1.84.
- [20] G. Zidane Dhamara and A. Nugroho, “Klasifikasi Genre Musik Menggunakan Machine Learning,” *Bulletin of Information Technology (BIT)*, vol. 6, no. 3, pp. 206–217, 2025, doi: 10.47065/bit.v5i2.2021.
- [21] S. Pouyanfar *et al.*, “A Survey on Deep Learning: Algorithms, Techniques, and Applications,” *ACM Comput. Surv.*, vol. 51, no. 5, Sep. 2018, doi: 10.1145/3234150.