

AUTOMASI PEMBANGKIT LIRIK LAGU DALAM BAHASA INDONESIA: IMPLEMENTASI GPT-NEO DALAM PEMROSESAN BAHASA KREATIF

Dino Hariatma Putra ¹, Naufal A. A. N. Syam ²

^{1,2} *Teknik Informatika/ Institut Teknologi Indonesia
Tangerang Selatan*

Email: ¹ dino.hp@iti.ac.id., ² azwir763@gmail.com

ABSTRAK

Kemajuan dalam pemrosesan bahasa alami (NLP) dan model pembelajaran mesin telah membuka peluang baru dalam pembuatan teks kreatif, termasuk lirik lagu. Penelitian ini mengevaluasi kemampuan model Generative Pre-trained Transformer (GPT-Neo) versi kecil (EleutherAI/gpt-neo-125m) untuk menghasilkan lirik lagu Bahasa Indonesia secara otomatis. Model ini dilatih ulang (fine-tuning) menggunakan dataset lirik lagu Indonesia yang telah diproses, dengan konfigurasi batch size 8, learning rate 5e-5, dan 50 epoch. Evaluasi dilakukan menggunakan metrik kuantitatif, termasuk BLEU score untuk menilai kelancaran dan koherensi teks. Hasil menunjukkan bahwa model mampu menghasilkan lirik yang koheren, relevan, dan menyerupai pola linguistik lirik Indonesia, dengan rata-rata BLEU score sebesar 0.72. Namun, model menghadapi tantangan dalam menjaga konsistensi tema pada lirik yang lebih panjang dan menangani variasi input seperti inkonsistensi tata bahasa. Meskipun demikian, hasil ini menunjukkan potensi besar GPT-Neo sebagai alat pembangkit teks kreatif yang relevan secara budaya. Penelitian ini memberikan dasar kuat untuk pengembangan lebih lanjut, terutama dalam meningkatkan kemampuan model untuk menangkap kompleksitas emosi dan tema dalam lirik lagu.

Kata Kunci: GPT-Neo, EleutherAI/gpt-neo-125m, lirik lagu Indonesia, pembangkitan teks, model bahasa, evaluasi BLEU.

ABSTRACT

Advancements in natural language processing (NLP) and machine learning models have opened new opportunities for generating creative texts, including song lyrics. This study evaluates the capability of the compact Generative Pre-trained Transformer model (GPT-Neo, EleutherAI/gpt-neo-125m) to generate Indonesian song lyrics autonomously. The model was fine-tuned using a curated dataset of Indonesian song lyrics, with configurations including a batch size of 8, a learning rate of 5e-5, and 50 epochs. The evaluation employed quantitative metrics, such as BLEU score to assess fluency and coherence. Results indicate that the model can produce coherent, contextually relevant lyrics that emulate the linguistic patterns of Indonesian songs, achieving an average BLEU score of 0.72. However, the model faces challenges in maintaining thematic consistency in longer lyrics and handling input variations, such as grammatical inconsistencies. Nevertheless, these findings highlight GPT-Neo's significant potential as a culturally relevant creative text-generation tool. This study provides a solid foundation for further development, particularly in enhancing the model's ability to capture emotional depth and thematic complexity in song lyrics.

Keywords: GPT-Neo, EleutherAI/gpt-neo-125m, Indonesian song lyrics, text generation, language model, BLEU evaluation.

I. PENDAHULUAN

Kemajuan terkini dalam pembelajaran mesin dan pemrosesan bahasa alami (NLP) telah secara signifikan meningkatkan kemampuan model bahasa generatif. Di antara model tersebut, Generative Pre-trained Transformers (GPT) telah mendapatkan pengakuan atas kemampuannya menghasilkan teks di berbagai domain yang logis dan sesuai dengan konteks. Salah satu perkembangan terbaru dalam bidang ini adalah GPT-Neo (EleutherAI/gpt-neo-125m) yang dikembangkan oleh EleutherAI, sebuah model open-source yang memberikan alat bagi akademisi dan pengembang untuk menghasilkan teks dalam jumlah besar yang tampak alami seperti buatan manusia. Meskipun penggunaan model GPT secara luas ditemukan dalam bahasa Inggris, sedikit penelitian yang mengkaji efektivitasnya dalam bahasa non-Inggris, termasuk Bahasa Indonesia, terutama dalam domain kreatif seperti penulisan lagu.

Keanekaragaman bahasa dan budaya dalam musik Indonesia menawarkan peluang sekaligus tantangan unik untuk pembuatan teks. Lirik lagu membutuhkan model yang mampu menghasilkan bahasa puitis sambil mempertahankan koherensi tematik dan dampak emosional, yang berbeda dengan domain teks terstruktur. Oleh karena itu, penting untuk menanyakan seberapa efektif model ini dapat beradaptasi dengan aktivitas kreatif dalam bahasa dan konteks yang berbeda dari data pelatihan aslinya ketika menggunakan GPT-Neo untuk menghasilkan lirik dalam Bahasa Indonesia.

Untuk menghasilkan lirik lagu dalam Bahasa Indonesia, penelitian ini bertujuan mengevaluasi versi ringkas dari model GPT-Neo. Kami meneliti kemampuan GPT-Neo dalam menghasilkan lirik yang logis, relevan secara budaya, dan ekspresif secara kreatif dengan melatih ulang model menggunakan dataset lirik lagu Indonesia. Penelitian ini mengevaluasi lirik yang dihasilkan berdasarkan orisinalitas, kelancaran, dan koherensi tematik, mengungkapkan kekuatan dan kelemahan model dalam konteks kreatif di luar tugas berbasis teks standar.

Temuan dari penelitian ini berkontribusi pada diskusi yang lebih luas tentang peran AI dalam seni dan memberikan wawasan penting tentang penggunaan model bahasa generatif dalam industri musik Indonesia. Selain itu, penelitian ini menyoroti potensi GPT-Neo sebagai alat penghasil teks dan parafrase yang dapat diterapkan di berbagai konteks kreatif, edukasi, dan komersial di Indonesia maupun luar negeri.

II. METODELOGI PENELITIAN

Dalam penelitian ini, model bahasa generatif open-source GPT-Neo, yang dikembangkan oleh EleutherAI,

digunakan untuk menghasilkan lirik lagu dalam Bahasa Indonesia. Metode yang digunakan mengevaluasi kelayakan GPT-Neo untuk menghasilkan teks kreatif melalui integrasi antara persiapan data, pelatihan model, dan evaluasi.

A. Pengumpulan dan Praproses Data

Pengumpulan Data: Kata-kata dari lirik lagu Indonesia dikumpulkan dari sumber yang tersedia secara publik di situs web kapanlagi.com. Dataset ini dirancang untuk mencakup berbagai gaya musik dan tema, sehingga memberikan dasar yang komprehensif untuk melatih model. Bagian pralatih menggunakan dataset Indonesian-nlp/wikipedia-id yang tersedia di huggingface.co.

Praproses Data: Elemen-elemen duplikat, simbol khusus, dan detail yang tidak relevan dihapus dari dataset. Untuk menjaga konsistensi tematik, lirik dibagi menjadi bagian-bagian yang sesuai, termasuk bait dan paduan suara. Tokenisasi digunakan untuk mengonversi data yang telah diproses menjadi urutan yang sesuai dengan format input model.

B. Gambaran Model

Pemilihan Model: GPT-Neo (EleutherAI/gpt-neo-125m) dipilih karena efisiensinya dalam komputasi dan kompatibilitasnya dengan dataset berukuran sedang. Model ini didasarkan pada arsitektur Transformer, yang menggunakan mekanisme perhatian-diri (self-attention) untuk memprediksi token berikutnya dalam suatu urutan.

Pre-trained: GPT-Neo dilatih sebelumnya menggunakan The Pile, sebuah dataset 800GB yang beragam dan berisi teks dari berbagai domain, sehingga melengkapinya dengan kemampuan linguistik yang luas. Tahap pralatih ini memberikan dasar yang kokoh untuk penyesuaian lebih lanjut menggunakan dataset lirik.

C. Pelatihan Ulang (Fine-Tuning)

Transfer Learning: Pelatihan ulang melibatkan penyesuaian bobot pralatih dari GPT-Neo ke dataset lirik lagu Indonesia. Langkah ini memanfaatkan prinsip transfer learning untuk meningkatkan performa model pada tugas-tugas khusus domain.

Proses Pelatihan: Model dilatih ulang menggunakan lingkungan yang dipercepat GPU. Hiperparameter seperti laju pembelajaran, ukuran batch, dan jumlah epoch dioptimalkan untuk menyeimbangkan kecepatan konvergensi dan kualitas output. Sebuah cosine decay learning rate scheduler diterapkan untuk memastikan stabilitas selama pelatihan.

D. Metode Evaluasi

BLEU (*Bilingual Evaluation Understudy*) score adalah metrik evaluasi yang umum digunakan dalam pembangkit bahasa (*text generation*) untuk mengukur kualitas teks yang dihasilkan oleh model dibandingkan dengan referensi atau target teks. BLEU score mengevaluasi seberapa mirip teks hasil model dengan referensi, menggunakan n-gram sebagai unit evaluasi.

III. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengevaluasi kemampuan GPT-Neo dalam menghasilkan lirik lagu Bahasa Indonesia melalui tahap pralatih dan pelatihan ulang (*fine-tuning*) pada dataset yang telah dikurasi. Pada tahap pralatih, model dikonfigurasi dengan ukuran batch 8 dan dilatih selama 3 *epoch* menggunakan sebagian dataset *Indonesian Parquet*. Nilai *loss* akhir yang dicapai adalah 0,771500, menunjukkan pembelajaran pola linguistik yang efektif. Proses pralatih berlangsung sekitar 70 menit dengan 18.720 iterasi, menunjukkan efisiensi model dalam mempelajari struktur bahasa Indonesia meskipun pelatihan awalnya berbasis data Bahasa Inggris.

Pelatihan ulang lebih lanjut meningkatkan kemampuan model untuk menghasilkan lirik yang sesuai dengan konteks. Konfigurasi utama termasuk ukuran batch 8, 50 *epoch*, laju pembelajaran (*learning rate*) sebesar 5e-5, *weight decay* sebesar 0,01, dan 500 langkah *warmup*. Model menyimpan keluaran *checkpoint* setiap 1.000 iterasi untuk memungkinkan evaluasi bertahap. Pelatihan ulang meningkatkan kemampuan model dalam menangkap tema, struktur, dan emosi lirik, sehingga sesuai dengan karakteristik lirik lagu Indonesia yang diinginkan. Namun, meskipun ada peningkatan ini, model masih menghadapi tantangan dalam menangani variasi input yang kompleks, seperti salah ketik, inkonsistensi kapitalisasi, dan kesalahan spasi.

A. Evaluasi Kuantitatif:

Tahap pralatih dan pelatihan ulang berhasil mengurangi *loss* model secara signifikan, menunjukkan kemampuan model yang semakin baik dalam memprediksi keluaran yang akurat. Selama pengujian, model yang telah dilatih ulang menghasilkan lirik yang koheren dan sesuai dengan tema, meskipun masih terdapat masalah dengan koherensi teks yang lebih panjang dan penanganan kesalahan yang kompleks.

B. Lirik yang dihasilkan

Prompt:

Tulis lirik lagu tentang perjalanan cinta penuh makna

Hasil:

*Malam-malam yang indah
Rindu-rindu lama
Paru suatu hari
Namun malu ku rasa apa yang tulus
Namun malu ku rasa tiada hanya
Atau pun mengerti
Janji untukku, bahagia mengalir
Cinta yang suci (Oh oh oh)
Cinta yang suci
Janji untukku, bahagia mengalir
Cinta yang suci
Janji untukku, bahagia mengalir
Oh oh oh*

C. Analisis

Lirik-lirik tersebut mengungkapkan tema cinta dan kesucian, tetapi menunjukkan ketidakkonsistenan sintaksis dan semantik pada beberapa bagian. Frase seperti "Paru suatu hari" mencerminkan masalah dengan koherensi, yang kemungkinan berasal dari

Metrik	Nilai	Interpretasi
<i>Pretraining Loss</i>	0.771500	Pembelajaran bahasa Indonesia yang efektif
<i>Fine-Tuning Epochs</i>	50	Pemahaman yang mendalam terhadap data lirik
<i>BLEU Score</i>	0.72	Kefasihan dan koherensi yang cukup

kurangnya paparan terhadap keberagaman tema selama proses pelatihan.

Table 1. Performance Metrics Table

KESIMPULAN

Penelitian ini menyoroti kemampuan GPT-Neo dalam menghasilkan lirik lagu Indonesia, menunjukkan kekuatan model ini sekaligus mengidentifikasi area yang membutuhkan peningkatan. Tahap pralatih menunjukkan bahwa GPT-Neo mampu memahami pola linguistik dalam bahasa Indonesia secara efektif, dengan nilai *loss* sebesar 0,771500, yang menunjukkan pelatihan dasar yang kuat. Penyempurnaan model dengan dataset yang dipilih dengan cermat meningkatkan kemampuannya untuk menghasilkan lirik yang koheren dan sesuai dengan konteks, mencerminkan kekayaan emosional dan koherensi tematik. Kinerja model, yang dinilai menggunakan metrik numerik *BLEU score* dengan nilai yang didapat adalah 0.72.

Nilai BLEU sebesar 0.72 dalam penelitian ini menunjukkan bahwa teks yang dihasilkan oleh model GPT-Neo memiliki tingkat kesesuaian yang cukup baik dengan teks referensi dalam dataset pelatihan. Ini berarti model mampu menghasilkan lirik yang koheren secara linguistik dan relevan dengan pola dan tema yang ada dalam data pelatihan, meskipun tidak selalu identik dengan referensi. **Kelancaran dan Koherensi Linguistik:** nilai ini menunjukkan bahwa struktur kalimat yang dihasilkan oleh model cukup padu dan menyerupai gaya bahasa lirik lagu Indonesia. Model berhasil menangkap pola n-gram (unigram, bigram, trigram, dll.) yang umum dalam data pelatihan. **Originalitas:** Meskipun model mengikuti pola data pelatihan, nilai BLEU di bawah 1.0 mengindikasikan bahwa teks yang dihasilkan tidak sepenuhnya identik dengan referensi. Ini berarti model memiliki kemampuan untuk menghasilkan lirik yang baru dan unik sambil tetap menjaga kesesuaian tema. **Kesesuaian Tema dan Konteks:** Nilai 0.72 menegaskan bahwa model mampu menghasilkan lirik yang sesuai dengan konteks prompt yang diberikan, seperti tema cinta atau perjalanan emosional, meskipun ada beberapa ketidaksempurnaan dalam koherensi di level semantik.

Namun, penelitian ini juga mengakui beberapa keterbatasan. Meskipun ada peningkatan yang dicapai melalui pelatihan ulang, model masih kesulitan dalam menangani variasi input yang rumit, seperti salah ketik dan ketidakkonsistenan kapitalisasi serta kesalahan spasi. Selain itu, mempertahankan koherensi tematik dan sintaksis dalam karya lirik yang lebih panjang masih menjadi tantangan. Tantangan ini menunjukkan bahwa meskipun GPT-Neo memiliki potensi besar untuk menghasilkan teks kreatif, penyempurnaan lebih lanjut diperlukan untuk mengatasi keterbatasan ini.

Secara umum, GPT-Neo menunjukkan janji yang besar dalam pembuatan otomatis lirik lagu Indonesia, terutama untuk tugas yang memerlukan kekayaan emosional dan tematik. Melalui pelatihan ulang tambahan dan modifikasi pada model, terutama dalam menangani inkonsistensi teks dan meningkatkan keberagaman tematik, model ini dapat menjadi aset berharga bagi sektor kreatif di Indonesia dan di tempat lain. Penelitian mendatang seharusnya fokus pada meningkatkan kapasitas model untuk mengelola berbagai format lirik, memperbaiki manajemen kesalahan, dan meningkatkan keterampilan generalisasi untuk menghasilkan lirik yang berkualitas dan relevan secara budaya di berbagai genre.

UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada Pusat Riset dan Pengabdian Masyarakat ITI telah memberikan hibah dalam penelitian ini.

REFERENSI

- [1] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, and D. Amodei, "Language Models are Few-Shot Learners," OpenAI, 2020. [Online]. Available: <https://arxiv.org/abs/2005.14165>.
- [2] Y.-Y. Chou, H.-T. Lin, and T.-L. Liu, "Adaptive and Generative Zero-Shot Learning," 2021. [Online]. Available: <https://arxiv.org/abs/2105.06136>.
- [3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Google AI Language, 2019. [Online]. Available: <https://arxiv.org/abs/1810.04805>.
- [4] P. Farhad, A. Moloud, Y. Luo, X. Zhou, C. P. Lim, X.-Z. Wang, and Q. J. Wu, "A Review of Generalized Zero-Shot Learning Methods," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 4051–4070, 2023. doi: 10.1109/TPAMI.2023.3221245.
- [5] "Generative AI Models Explained," AltexSoft, October 13, 2022. [Online]. Available: <https://www.altexsoft.com/blog/generative-ai/>.
- [6] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, and Y. Bengio, "Generative Adversarial Nets," 2014. [Online]. Available: <https://arxiv.org/abs/1406.2661>.
- [7] T. He, X. Tan, Y. Xia, D. He, T. Qin, Z. Chen, and T.-Y. Liu, "Layer-Wise Coordination between Encoder and Decoder for Neural Machine Translation," 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), 2018. [Online]. Available: <https://arxiv.org/abs/1809.00182>.
- [8] T. Iqbal and S. Qureshi, "The Survey: Text Generation Models in Deep Learning," Journal of King Saud University – Computer and Information Sciences, pp. 2516–2526, 2020. doi: 10.1016/j.jksuci.2020.01.003.
- [9] A. Lamb, "A Brief Introduction to Generative Models," Department of Informatics and Operations Research, 2021. [Online]. Available: <https://arxiv.org/abs/2106.00919>.
- [10] D. T. Larose and C. D. Larose, *Discovering Knowledge in Data*, Wiley, 2014.
- [11] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, and D. Kiela, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," 2021. [Online]. Available: <https://arxiv.org/abs/2005.11401>.
- [12] E. D. Liddy, *Natural Language Processing*, Marcel Dekker, Inc., 2001.

-
-
- [13] Z. Liu, W. Lin, Y. Shi, and J. Zhao, "A Robustly Optimized BERT Pre-training Approach with Post-training," 2021, pp. 1218–1227. [Online]. Available: <https://arxiv.org/abs/1907.11692>.
- [14] R. Luo, L. Sun, Y. Xia, T. Qin, S. Zhang, H. Poon, and T.-Y. Liu, "BioGPT: Generative Pre-trained Transformer for Biomedical Text Generation and Mining," 2023. [Online]. Available: <https://arxiv.org/abs/2210.10341>.
- [15] B. Mahesh, "Machine Learning Algorithms - A Review," *International Journal of Science and Research (IJSR)*, pp. 381–386, 2020. doi: 10.21275/SR202011112224.
- [16] A. Musyafa, Y. Gao, A. Solyman, C. Wu, and S. Khan, "Automatic Correction of Indonesian Grammatical Errors Based on Transformer," *Applied Science*, 2022. doi: 10.3390/app12041987.
- [17] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving Language Understanding," 2018. [Online]. Available: <https://www.cs.cmu.edu/~rsalakhu/papers/lm-dnn.pdf>.
- [18] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language Models are Unsupervised Multitask Learners," OpenAI, 2019. [Online]. Available: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multi_task_learners.pdf.
- [19] S. Rothe, S. Narayan, and A. Severyn, "Leveraging Pre-trained Checkpoints for Sequence Generation Tasks," 2020. [Online]. Available: <https://arxiv.org/abs/1907.12461>.
- [20] A. Vaswani et al., "Tensor2Tensor for Neural Machine Translation," 2018. [Online]. Available: <https://arxiv.org/abs/1803.07416>.
- [21] A. Vaswani et al., "Attention Is All You Need," 31st Conference on Neural Information Processing Systems (NIPS 2017), 2017. [Online]. Available: <https://arxiv.org/abs/1706.03762>.
- [22] T. Wolf et al., "Huggingface Transformers: State-of-the-Art Natural Language Processing," 2020, pp. 38–45. [Online]. Available: <https://arxiv.org/abs/1910.03771>. M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in *Proc. ECOC'00*, 2000, paper 11.3.4, p. 109.
- [23] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.
- [24] (2002) The IEEE website. [Online]. Available: <http://www.ieee.org/>
- [25] M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: <http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/>