

# Generator Kisah Daerah Berbasis Bahasa Jawa dengan Finetuned GPT-2

Kayla Queenazima Santoso<sup>1</sup>, Karunia Perjuangan Mustadl'afin<sup>2</sup>, Lutfi Andriyanto<sup>3</sup>, Igi Ardiyanto<sup>4</sup>\*

<sup>1</sup>Fakultas Teknik Universitas Gadjah Mada, Yogyakarta, 55281, Indonesia, kaylaqueenazima@mail.ugm.ac.id

<sup>2</sup>Fakultas Teknik Universitas Gadjah Mada, Yogyakarta, 55281, Indonesia, karuniaperjuangan@mail.ugm.ac.id

<sup>3</sup>Fakultas Teknik Universitas Gadjah Mada, Yogyakarta, 55281, Indonesia, lutfiandriyanto@mail.ugm.ac.id

<sup>4</sup>Fakultas Teknik Universitas Gadjah Mada, Yogyakarta, 55281, Indonesia, igi@ugm.ac.id

**INTISARI** — Pengembangan *text generator* pada Bahasa Jawa bertujuan menciptakan generator kisah daerah yang memanfaatkan dua skema dataset, yaitu data hasil translasi dan data teks/dokumen digital berbahasa Jawa dengan variasi dua model *pre-trained* GPT-2. Pendekatan ini dimaksudkan untuk mengatasi masalah kualitas dan kehilangan unsur karakter kultural serta linguistik dalam proses pengumpulan dataset. Dengan demikian, pengembangan ini diharapkan dapat meningkatkan ekonomi dan pelestarian budaya Indonesia melalui kreativitas *storytelling* berbahasa Jawa. Metode yang digunakan ialah *fine-tuning* GPT-2 *pre-trained*. Dari pengembangan metode tersebut, didapatkan skema model *finetuned* GPT-2 Medium Berbahasa Indonesia dengan skema dataset II serta *scheduler inverse square root* dengan perplexity 65,70.

**KATA KUNCI** — *Text Generator* Berbahasa Jawa, GPT-2, *Large Language Model*, *Low-Resource Language Model*

## I. LATAR BELAKANG

Budaya beragam Indonesia membawa cerita pada tiap bahasa, tarian, motif, adat istiadat, dan dimensi budaya lainnya sehingga menjadi nilai autentik pengenalan budaya Indonesia. Kemampuan bercerita atau *storytelling* dalam perkembangannya berperan penting dalam meningkatkan nilai ekonomi serta pelestarian budaya melalui pengenalannya dalam berbagai produk baik benda (pakaian, perabot, dsb) hingga tak benda (lagu, festival, dsb). Salah satu festival tradisi besar Korea Selatan, Gangneung Danoje Festival, bahkan mengundang lebih dari 2 miliar pengunjung pada 2017. Nilai merek festival tersebut dikembangkan sedemikian rupa dengan membawa konten budaya lokal, memperhatikan atribut *storytelling*, dan keterlibatan warga lokal [1].

Bahasa merupakan salah satu aspek budaya yang menarik. Indonesia memiliki 719 bahasa daerah dengan 707 bahasa masih memiliki penutur aktif [2]. Berdasarkan sejarah, penggunaan bahasa daerah pernah menjadi bagian legendaris karya fenomenal 'Warkop DKI', Peningkatan kreativitas *storytelling* budaya Indonesia menjadi sangat penting untuk meningkatkan nilai ekonomi budaya pariwisata Indonesia sekaligus pelestarian budaya utamanya menuju Indonesia maju dan keberlanjutan.

Peluang potensi peningkatan ekonomi sekaligus pelestarian budaya Indonesia ini dapat diakselerasi dengan memberi perspektif baru melalui pengembangan generator kisah daerah berbasis kecerdasan buatan memanfaatkan pendekatan *text generator*. *Text generator* dapat memberi perspektif-perspektif baru penulisan cerita dan mengakselerasi proses pelestarian dengan akselerasi proses pembuatan bahan bacaan yang bervariasi.

Bahasa Jawa merupakan salah satu bahasa daerah yang paling mudah didapatkan datanya bahkan merupakan bahasa ke-26 yang paling banyak digunakan dengan 68 miliar penutur aktif di seluruh dunia [3]. Selain itu, Bahasa Jawa bersama dengan dialek, adat istiadat, tarian, lagu, cerita rakyat, dan dimensi budaya lainnya begitu kaya dan bervariasi pada tiap daerah yang berpeluang besar untuk dieksplorasi melalui *text generator*.

Ketersediaan data yang terbatas menjadi tantangan dalam pengembangan model *text generator* berbasis bahasa daerah karena jumlah penutur aktif atau dokumentasi yang terbatas. Berangkat dari problematika tersebut, metode yang diusulkan berupa pengembangan *text generator* pada bahasa daerah,

Bahasa Jawa, dengan tujuan mengoptimalkan *finetune pre-trained* model GPT-2 melalui perspektif variasi desain dataset, parameter model, dan basis bahasa *pre-trained*. Data yang digunakan terbatas pada teks cerita berdialek bahasa Jawa daerah Jawa Tengah dan DI Yogyakarta. Melalui pengujian metode tersebut, penelitian ini berpotensi bermanfaat sebagai gebrakan inovasi industri kreatif sekaligus upaya pelestarian bahasa daerah. Selain itu, penelitian ini berkontribusi dalam kajian pengembangan model bahasa dengan jumlah dataset terbatas melalui perspektif sumber dataset, jumlah parameter model, dan basis bahasa *pre-trained* model.

## II. KAJIAN PUSTAKA

### A. ETL PIPELINE

ETL (*Extraction, Transformation, Loading*) adalah proses mengumpulkan informasi dari berbagai sumber dan mengolahnya ke dalam sebuah data yang terstruktur dan homogen. Proses ini terdiri dari tiga bagian [4]. Pada bagian (1) *Extraction*, data yang ada dalam berbagai sumber dikumpulkan dalam satu tempat penyimpanan dengan format yang berbeda-beda. Kemudian bagian (2) *Transformation*, semua data dengan skema dan format yang berbeda dibersihkan dan diubah menjadi satu format konsisten. Terakhir, bagian (3) *Loading*, data yang sudah diubah disimpan ke dalam satu lokasi untuk dimanfaatkan.

### B. GPT

*Generative Pre-trained Transformer* atau GPT adalah salah satu model kecerdasan buatan untuk pemrosesan bahasa alami. Model ini ini bersifat umum dan agnostik terhadap tugas pengolahan bahasa alami sehingga dapat dikembangkan menjadi solusi permasalahan kebahasaan tertentu maupun umum, seperti melengkapi teks, menjawab pertanyaan, mendeteksi kemiripan makna, serta klasifikasi dokumen [5].

Model GPT merupakan model *semi-supervised learning*, artinya dapat dibuat dengan data tanpa label dan terlabel. Ada 2 tahapan proses pembelajaran GPT. Pertama, proses *unsupervised training*. Pada proses ini, data korpus bahasa tidak

berlabel diubah ke vektor, kemudian dilatih pada model atensi berbasis Transformer [6] sehingga didapat model bahasa umum. Kedua, dilakukan *supervised fine-tuning*, yaitu spesialisasi model GPT ke suatu tugas khusus seperti generasi teks cerita/merangkum dokumen, dengan menggunakan data berlabel berupa sekumpulan data masukan dan juga keluaran yang diharapkan.

**C. LLM (Large Language Model) untuk Bahasa Daerah**

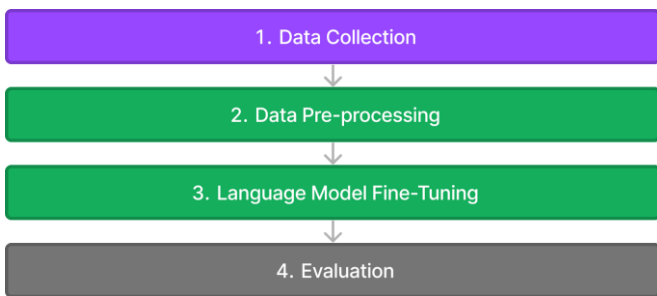
Pengembangan LLM berbasis bahasa daerah memiliki tantangan utama minimnya dokumentasi corpus yang dapat digunakan. Berdasarkan [7], pengembangan LLM dengan dataset minim dapat mengalami peningkatan performa dengan menggunakan *pre-trained* model berbasis bahasa Inggris. Hal ini dapat menjadi peluang pengembangan LLM bahasa baru berdataset minim dengan memanfaatkan *pre-trained* berbasis bahasa lain. Penelitian sejenis (Tabel I) berupa LLM berbahasa Jawa [8] membandingkan penggunaan *pre-trained* dari beberapa sumber seperti GPT-2, BERT, RoBERTa, dan DistilBert dengan tipe causal dan masked menggunakan dataset Bahasa Jawa berbasis wikipedia Bahasa Jawa.

**TABEL I**  
PERBANDINGAN PENELITIAN YANG DILAKUKAN DENGAN PENELITIAN SEBELUMNYA

Judul	Causal and Masked Language Modeling of Javanese Language using Transformer-based Architectures [8]	Penelitian ini
Model	<i>Pre-trained</i> Causal LM GPT-2 small, BERT, RoBERTa, dan DistilBert	<i>Pre-trained</i> Causal LM GPT-2 Medium Indonesia dan Small Jawa
Dataset	80,000 artikel wikipedia Jawa	707 artikel cerita Jawa
Hasil	<i>finetuned</i> pada task <i>downstream</i> dan klasifikasi sentimen	<i>finetuned</i> dengan 2 desain dataset

**III. METODE YANG DIUSULKAN**

Alur skema secara garis besar terdiri dari pengumpulan data, pemrosesan data, pelatihan model fine-tuning, dan evaluasi model seperti Gambar 1.

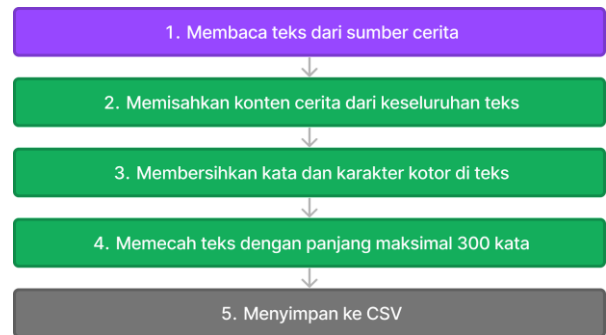


Gambar 1. Skema Besar Metode yang Diusulkan

**A. Data Collection**

Metode pengumpulan data yang penulis gunakan yakni menggunakan proses ETL (*Extract, Transform, and Load*). Penulis mengumpulkan dataset dari beberapa sumber sehingga terdapat 2 skema ETL yang menghasilkan 2 skema dataset.

1) Pengumpulan Cerita Skema I: Berbahasa Jawa

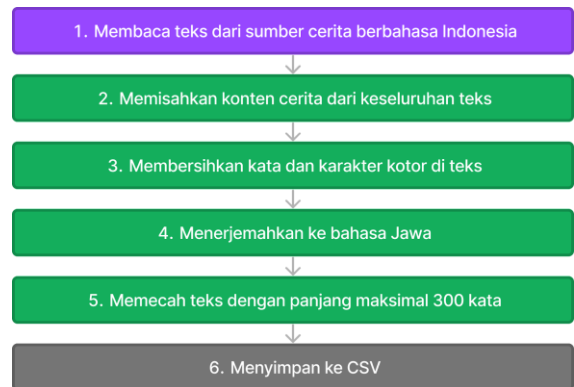


Gambar 2. Proses ETL Skema I: Cerita Berbahasa Jawa

Pada skema ini, dataset bersumber dari cerita dari website, Academia.edu, dan dokumen PDF dengan kata kunci cerita, babad, maupun naskah teater berbahasa Jawa.

Website dan dokumen memiliki proses ETL serupa (Gambar 2) dengan library yang berbeda. Tahap pertama adalah proses *extract* yang berisi pembacaan teks. Tahap kedua hingga keempat adalah proses *transform* sehingga didapat tiap baris teks maksimal 300 kata sesuai dengan batas masukan pada model GPT-2. Terakhir adalah proses *load* yang menyimpan kumpulan data ke dalam CSV.

Untuk dokumen PDF, penulis menggunakan library PyMuPDF yang bisa membaca karakteristik font dan warna background. Hal tersebut akan digunakan untuk pembersihan teks awal, yakni menghilangkan teks selain cerita seperti menghilangkan judul, nomor halaman, tulisan *header* dan *footer*, dan keterangan gambar. Sementara itu, untuk website, penulis menggunakan library *request* untuk mengambil teks HTML dan mengambil cerita dari teks tersebut tanpa mengambil tulisan judul, *navbar*, dan lain-lain yang ada di website dengan melakukan seleksi menggunakan BeautifulSoup.



Gambar 3. Proses ETL Skema II: Cerita Terjemahan Berbahasa Jawa

Proses ketiga (membersihkan kata dan karakter kotor di teks) di Gambar 2 maupun Gambar 3 menggunakan cara yang sama. Berikut pembersihan teks yang dilakukan.

- Mengubah ke karakter ASCII.

Mengubah semua karakter menjadi karakter ASCII. Contoh karakter yang diubah yakni mengubah karakter “é” menjadi “e” dengan library *unidecode*.

- Mengganti semua whitespace.

Mengganti karakter whitespace seperti spasi, enter, tab, dan lain-lain menjadi karakter spasi dengan *regex*.

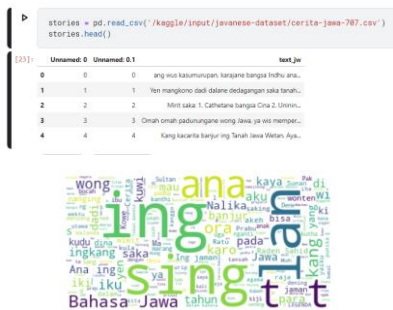
- Menghapus URL

Menghapus url dengan pola “<site>.<domain>” seperti cerita.id, cerita.com, dan lain-lain dengan regex.

- Filtering berdasar jumlah karakter

Menghapus baris dari teks yang memiliki karakter kurang dari 5 karakter untuk menghindari batas cerita seperti “\*\*\*”, “---”, dan lainnya.

Proses keempat pada Gambar 2 dan kelima pada Gambar 3 (memecah teks dengan maksimal 300 kata) merupakan proses memecah sekumpulan teks dari data *source* menjadi list dari kumpulan kalimat-kalimat yang maksimal berjumlah 300 kata. Agar konteks kalimat tetap utuh, pemisahan teks berdasarkan kalimat bukan kata. Selain itu, khusus untuk dari dokumen pdf, dilakukan juga pemisahan terhadap bab (*chapter*) agar konteks dari cerita tidak rancu. Tiap kumpulan kalimat yang dipisahkan akan menjadi satu datum yang direpresentasikan ke dalam satu baris di file CSV.



**Gambar 4.** Sampel Dataset (atas) dan Wordcloud (bawah) Skema I

Dari Skema I didapatkan dataset yang didominasi gaya teks cerita rakyat ditandai dengan munculnya kata-kata ‘sultan’, ‘prabu’, dan ‘ratu’ seperti pada Gambar 4.

2) Pengumpulan Cerita Skema II: Terjemahan Bahasa Jawa



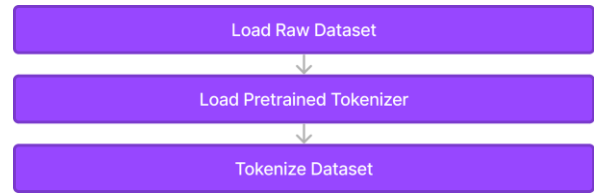
**Gambar 5.** Sampel Dataset (atas) dan Wordcloud (bawah) Skema II

Perbedaan skema II dengan skema I ialah penulis mengambil cerita berbahasa Indonesia dari website berisi cerita pendek dan menerjemahkannya ke dalam bahasa Jawa dengan library GoogleTrans.

Dari skema II didominasi gaya teks cerita anak-anak atau cerita populer dengan melihat adanya kata-kata ‘sekolah’ dan ‘aku’ seperti pada Gambar 5.

Hasil dari proses pengumpulan data ini adalah dua dataset, yakni dataset cerita berbahasa Jawa dan dataset cerita terjemahan berbahasa Jawa. Masing-masing dari dataset tersebut memiliki 707 baris.

**B. Data Pre-processing**

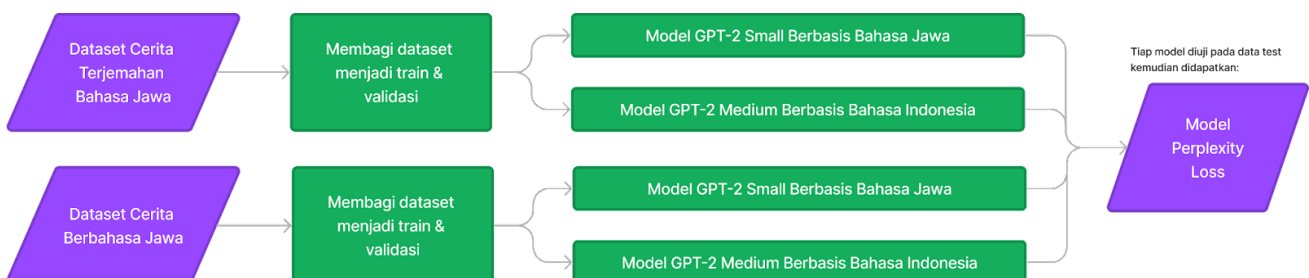


**Gambar 6.** Skema Data Pre-processing

Pada tahap ini (Gambar 6), data yang terdiri dari kumpulan kalimat atau teks diubah menjadi kumpulan vektor yang berfungsi sebagai masukan model. Untuk mengubahnya, terdapat tokenizer yang memetakan token ke vektor tersebut. Token tersebut adalah kumpulan huruf yang ada di teks. Tokenizer yang mengelompokkan beberapa huruf sebagai satu token tersebut di-train menggunakan Byte Pair Encoding (BPE). Dengan demikian, teks masukan untuk model akan berubah menjadi runtutan vektor. Setelah itu, runtutan vektor tersebut bisa langsung dimasukkan ke model yang bertugas memprediksi vektor token selanjutnya dari runtutan vektor yang ada. Lalu, token tersebut berubah menjadi satu atau beberapa huruf yang melengkapi teks. Proses prediksi tersebut berulang sampai panjang teks tertentu sudah terpenuhi atau muncul token khusus di prediksi, yaitu <endoftext>.

**C. Large Language Model Fine Tuning**

*Language model pre-trained* (Gambar 7) yang digunakan ialah GPT-2 *medium fine-tuned* bahasa Indonesia [9] dengan parameter 117M dan GPT-2 *small fine-tuned* bahasa Jawa [8] dengan parameter 345M. Dataset berupa 2 skema *data collection*. Library yang digunakan yaitu Transformers dan PyTorch. Tiap skema dataset diambil 50 baris lalu digabung sebagai data *test* kemudian 657 baris dibagi menjadi *train* dan validasi (9:1) dengan seed 456368. Proses pelatihan menggunakan Trainer dari Transformers diatur weight decay sebesar 0.05 pada semua layer kecuali bias dan LayerNorm pada optimizer, fungsi optimizer AdamW, batch size 1, learning rate 5e-5, 10 jumlah step untuk warmup linear, menggunakan 16-bit precision training, dan epoch 20. Hasil dari fine-tuning berupa



**Gambar 7.** Sampel Dataset (atas) dan Wordcloud (bawah) Skema I

model, data learning rate, perplexity terhadap data validasi, dan loss tiap 73 steps.

#### D. Evaluasi

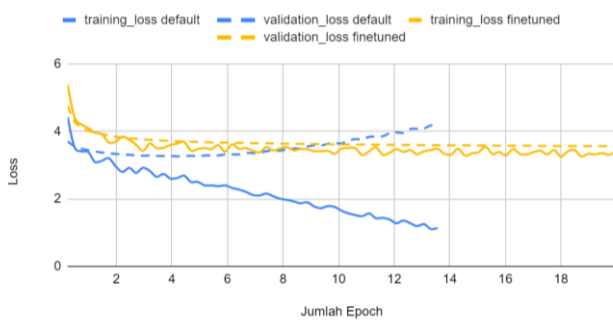
Evaluasi pada model dilakukan dengan 2 metrik yaitu loss menggunakan *multi cross entropy loss* (Persamaan 3.1) dan perplexity (Persamaan 3.2). Proses kalkulasi evaluasi dilakukan pada data test yang diambil dari 100 baris data kombinasi skema I dan II kemudian dibandingkan dengan model text generator berbasis Bahasa Jawa [9].

$$Loss = - \sum_{i=1}^N y_i \log(p_i) \quad (3.1)[10]$$

$$Perplexity = \exp \left\{ \frac{1}{N} \sum_{i=1}^N Loss_i \right\} \quad (3.2)[11]$$

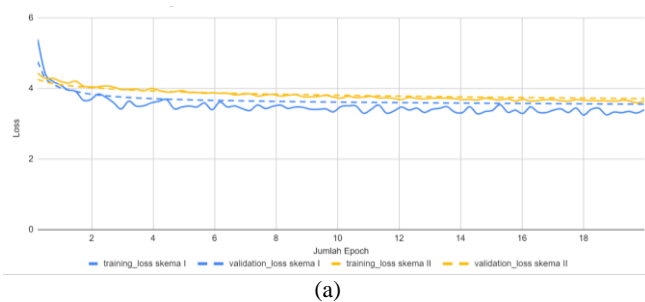
### IV. HASIL DAN ANALISIS

#### A. Finetuned LLM GPT-2 pada Dataset Bahasa Jawa

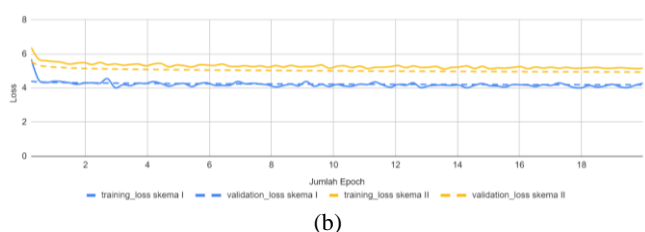


Gambar 8. Perbandingan Loss pada Model Sebelum dan Sesudah *Finetuned* Parameter

Pengembangan LLM pada Bahasa Jawa diujicoba terlebih dahulu dengan pelatihan menggunakan pre-trained IndonesiaNLP [9] default (scheduler learning rate linear) kemudian dibandingkan dengan penggunaan scheduler learning rate Inverse Square Root (Gambar 8). Pada Gambar 8 dapat dilihat model default mengalami overfitting di sekitar epoch 1 ditandai dengan jarak loss validasi terhadap train yang semakin besar, bahkan loss validasi justru meningkat. Hal ini



(a)



(b)

Gambar 9. Perbandingan Loss Model GPT-2 Medium Berbasis Bahasa Indonesia (a) dan Bahasa Jawa (b)

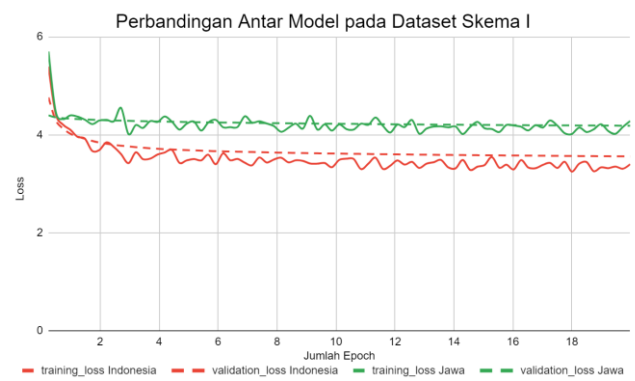
diperkirakan karena dataset yang sangat kecil sehingga model

mudah overfitting. Salah satu upaya mitigasinya ialah dengan menurunkan learning rate lebih awal. Terbukti dengan model finetuned, grafik loss pada Gambar 8 menunjukkan loss validasi dan train memiliki selisih yang kecil walaupun loss training pada model default lebih rendah daripada finetuned. Oleh karena itu, dapat disimpulkan bahwa penggunaan scheduler inverse square root berhasil mencegah overfitting pada pelatihan LLM.

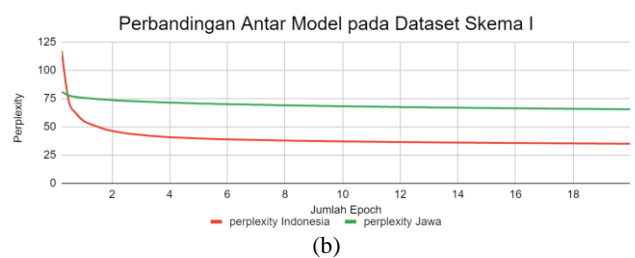
#### B. Data-Centric Approach

Pengaturan parameter melalui finetuned model kemudian dikembangkan untuk proses pelatihan 2 model pre-trained yaitu GPT-2 Medium berbasis Bahasa Indonesia [9] (Gambar 9 (a)) dan GPT-2 Small berbasis Bahasa Jawa [8] (Gambar 9 (b)). Pada kedua model, didapatkan loss yang tidak overfit maupun underfit serta grafik loss skema I dominan lebih rendah daripada skema II. Bahkan, pada GPT-2 small Jawa selisih antara skema I dan II lebih lebar daripada finetuned GPT-2 medium Indonesia. Hal ini diperkirakan karena kualitas dataset skema I berupa teks yang ditulis berbahasa Jawa memiliki kualitas dan validitas dataset yang lebih tinggi daripada teks hasil terjemahan (skema II). Oleh karena itu, pada jumlah dataset maupun epoch yang sama bahkan pada model yang berbeda, dataset skema I memiliki loss lebih rendah daripada skema II sehingga dapat disimpulkan dataset yang lebih minim memiliki performa lebih tinggi apabila validitas dataset tinggi walaupun tidak signifikan pada Gambar 4.2 (a).

#### C. Perbedaan Parameter Model dan Basis Bahasa



(a)



(b)

Gambar 10. Perbandingan Antar Model pada Dataset Skema I Berdasarkan Loss (a) dan Perplexity (b)

Melalui perbandingan antar skema dan finetuned, kedua model GPT-2 skema I kemudian dibandingkan loss dan perplexity tiap epoch. Berdasarkan Gambar 10 dapat dilihat model GPT-2 Medium Indonesia lebih rendah daripada GPT-2 Jawa baik loss maupun perplexity. Hal ini diperkirakan karena parameter model GPT-2 medium 2x lipat GPT-2 small walaupun GPT-2 small sudah berbasis Bahasa Jawa, sementara GPT-2 medium berbasis Bahasa Indonesia. Selain itu, nilai loss

dan perplexity tiap epoch turun sedikit demi sedikit disebabkan scheduler learning rate inverse square root yang digunakan sehingga learning rate sangat rendah, namun penurunan menunjukkan potensi pengembangan performa model melalui peningkatan jumlah epoch. Oleh karena itu, dapat disimpulkan bahwa pada dataset berukuran minim, penggunaan pre-trained model dengan akar bahasa yang mirip dan parameter besar lebih berdampak daripada parameter sedikit namun bahasa yang sama.

Dari keseluruhan ujicoba yang dilakukan, hasil evaluasi tiap

**TABEL II**  
LOSS DAN PERPLEXITY TIAP SKEMA PADA DATA TEST

GPT-2 Medium Bahasa Indonesia			GPT-2 Small Bahasa Jawa		
Skema	Loss	Perplexity	Skema	Loss	Perplexity
Default	4,859	128,961	Pre-Trained	6,140	464,490
Dataset I	4,286	72,717	Dataset I	5,749	313,926
Dataset II	4,185	65,70	Dataset II	5,703	300,060

model terhadap data test (Tabel II) menunjukkan loss dan perplexity paling rendah pada finetuned pre-trained GPT-2 Medium Bahasa Indonesia. Pemilihan skema dataset I dan II tidak berpengaruh signifikan dan konsisten pada proses evaluasi, namun terdapat selisih tipis dengan skema dataset I lebih rendah pada proses pelatihan terutama pelatihan maupun evaluasi GPT-2 Small Bahasa Jawa.

## V. KESIMPULAN

Dengan dua dataset yang berbeda, penulis mengembangkan empat skema model. Skema I menggunakan dataset cerita bahasa Jawa dan Skema II menggunakan dataset cerita bahasa Jawa hasil terjemahan. Kedua skema tersebut dilatih pada pre-trained GPT-2 Medium Berbahasa Indonesia [9] dan GPT-2 Small Berbahasa Jawa [8]. Dari proses finetuned didapatkan model memiliki performa lebih baik berupa loss (4.185) dan perplexity (65.70) paling rendah serta tidak overfit pada model finetuned GPT-2 Medium Berbahasa Indonesia dengan skema dataset II serta scheduler inverse square root. Namun, dalam proses pelatihan, terdapat selisih kecil dataset I dan II sehingga validitas dataset menjadi penting dalam pertimbangan peningkatan performa walaupun tidak signifikan. Oleh karena itu, dapat disimpulkan performa pada text generator LLM Bahasa Jawa dapat ditingkatkan performanya melalui peningkatan parameter walaupun basis bahasa berbeda asalkan masih memiliki akar bahasa yang banyak beririsan serta peningkatan validitas dataset walaupun pada model dengan parameter besar tidak berdampak signifikan.

## REFERENSI

[1] 이계용, J., & Yoo, K. W. (2018). The structural relationship between selection attributes of traditional culture festival storytelling, brand equity, brand attitude, and brand loyalty. *The Journal of the Korea Contents Association*, 18, 647-659.

[2] Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, "Badan Bahasa petakan 652 bahasa daerah di Indonesia," Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, [Online]. Available: <https://www.kemdikbud.go.id/main/blog/2018/07/badan-bahasa-petakan-652-bahasa-daerah-di-indonesia>. [Accessed: Jul. 1, 2023].

[3] D. M. Eberhard, G. F. Simons, and C. D. Fennig (eds.), "Ethnologue: Languages of the World," 26th ed. Dallas, Texas: SIL International, 2023. [Online]. Available: <http://www.ethnologue.com>.

[4] R. Wijaya and B. Pudjoatmodjo, "An overview and implementation of extraction-transformation-loading (ETL) process in data warehouse (Case

study: Department of agriculture)," 2015 3rd International Conference on Information and Communication Technology (ICoICT), Nusa Dua, Bali, Indonesia, 2015, pp. 70-74, doi: 10.1109/ICoICT.2015.7231399.

[5] Radford, A., Narasimhan, K., Salimans, T., dan Sutskever, I. 2018. Improving Language Understanding by Generative Pre-Training. OpenAI.

[6] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., dan Polosukhin, I. 2017. Attention is all you need. In *Advances in neural information processing systems*. pp.5998-6008.

[7] Z. Chi, H. Huang, L. Liu, Y. Bai, X. Gao and X. -L. Mao, "Can Pretrained English Language Models Benefit Non-English NLP Systems in Low-Resource Scenarios?," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, doi: 10.1109/TASLP.2023.3267618.

[8] W. Wongso, D. S. Setiawan and D. Suhartono, "Causal and Masked Language Modeling of Javanese Language using Transformer-based Architectures," 2021 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Depok, Indonesia, 2021, pp. 1-7, doi: 10.1109/ICACSIS53237.2021.9631331.

[9] Akmal et al., "Indonesian-NLP/GPT2-medium-Indonesian · hugging face," *indonesian-nlp/gpt2-medium-indonesian · Hugging Face*, <https://huggingface.co/indonesian-nlp/gpt2-medium-indonesian> (accessed Jul. 1, 2023).

[10] I. Zafar and G. Tzanidou, *Hands-on Convolutional Neural Networks with Tensorflow: Solve Computer Vision Problems with Modeling in Tensorflow and Python*. Birmingham ; Mumbai: Packt, 2018.

[11] Perplexity of fixed-length models, <https://huggingface.co/docs/transformers/perplexity> (accessed Jun. 30, 2023).