

ABSTRAK

Penelitian ini bertujuan untuk mengimplementasikan model *Generative Pre-trained Transformers* (GPT) sebagai generator pantun dalam bahasa Indonesia. Permasalahan yang dihadapi dalam pembuatan pantun berbahasa Indonesia meliputi keterbatasan pengetahuan kata, variasi yang terbatas, dan sulitnya menjaga konsistensi pola dan sajak. Untuk mengatasi permasalahan ini, penelitian ini mengusulkan penggunaan model GPT sebagai pantun generator. Melalui metode yang mencakup tahap *preprocessing dataset* dan *fine-tuning* menggunakan dataset pantun berbahasa Indonesia, model GPT berhasil mencapai hasil akhir dengan *perplexity* sebesar 16.941, menunjukkan tingkat akurasi yang tinggi dalam memprediksi teks. Meskipun belum berhasil menghasilkan pantun dengan pola sajak yang sempurna, model GPT tetap mampu menghasilkan pantun-pantun dengan struktur dan karakteristik yang konsisten dengan pantun berbahasa Indonesia. Dalam pengujian dan implementasi ke dalam streamlit, model GPT pantun generator mampu menghasilkan pantun-pantun kreatif dan dapat disesuaikan dengan input pengguna. Dengan menggabungkan kecerdasan buatan dan tradisi sastra Indonesia, aplikasi GPT pantun generator ini memiliki potensi untuk membantu mempromosikan dan melestarikan kebudayaan Indonesia, sambil memberikan hiburan, edukasi, dan potensi dalam industri kreatif.

Kata Kunci: Bahasa Indonesia, *Fine-Tuning*, *Generative Pre-trained Transformers* (GPT), Generator Pantun, *Preprocessing*

ABSTRACT

This study aims to implement the Generative Pre-trained Transformers (GPT) model as a pantun generator in the Indonesian language. The problems faced in creating Indonesian pantun include limited vocabulary knowledge, limited variations, and difficulties in maintaining consistent patterns and rhyme. To address these issues, this study proposes the use of the GPT model as a pantun generator. Through a method that encompasses preprocessing the dataset and fine-tuning using Indonesian pantun datasets, the GPT model achieved a final perplexity of 16.941, indicating a high level of accuracy in text prediction. Although the model has not yet achieved perfect rhyme patterns, it is still capable of generating pantun with consistent structure and characteristics of Indonesian pantun. In testing and implementation into Streamlit, the GPT pantun generator model is capable of producing creative pantun that can be customized based on user input. By combining artificial intelligence and Indonesian literary traditions, this GPT pantun generator application has the potential to help promote and preserve Indonesian culture while providing entertainment, education, and opportunities in the creative industry.

Keywords: *Fine-Tuning, Indonesian Language, Generative Pre-trained Transformers (GPT), Pantun Generator, Preprocessing*

DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS	ii
HALAMAN PENGESAHAN	iii
KATA PENGANTAR	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR / SKRIPSI UNTUK KEPENTINGAN AKADEMIS	vi
ABSTRAK.....	vii
<i>ABSTRACT</i>	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL.....	xiii
BAB 1 PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Tujuan Penelitian.....	2
1.4. Ruang Lingkup.....	3
1.5. <i>State of The Art</i>	3
1.6. Sistematika Penulisan.....	8
BAB 2 LANDASAN TEORI.....	9
2.1. <i>Machine Learning</i>	9
2.2. <i>Deep Learning</i> (Ilahiyah & Nilogiri, 2018)	10
2.3. <i>Preprocessing Data</i> (Tineges & Davita, 2021).....	11
2.4. <i>Transformers</i> (Vaswani, dkk., 2017)	11
2.4.1. <i>Encoder dan Decoder</i>	13
2.4.2. <i>Attention</i>	14

2.5.	<i>Generative Pre-Trained Transformers</i>	15
2.5.1.	<i>Generative Pre-Trained Transformers 1</i> (Radford, dkk., 2018).....	16
2.5.2.	<i>Generative Pre-Trained Transformers 2</i> (Radford, dkk., 2019).....	16
2.5.3.	<i>Generative Pre-Trained Transformers 3</i> (Brown, dkk., 2020).....	17
2.6.	<i>PyTorch</i> (Tran, 2020).....	17
2.7.	<i>Streamlit</i> (Singh, 2021)	18
2.8.	<i>Natural Language Processing</i> (Liddy, 2001)	18
BAB 3 METODOLOGI PENELITIAN		20
3.1.	Tahapan Penelitian	20
3.2.	Identifikasi Masalah	22
3.3.	Studi Literatur	24
3.4.	Pengumpulan Data	24
3.5.	<i>Preprocessing</i>	27
3.6.	<i>Pre-Training</i>	35
3.7.	<i>Fine-Tuning</i>	36
3.7.1.	Spesifikasi <i>Hardware</i> dan <i>Software Environment</i>	37
3.7.2.	<i>Tuning Hyperparameter</i>	38
3.7.3.	<i>Modelling Pantun</i>	39
BAB 4 HASIL DAN PEMBAHASAN		40
4.1.	Evaluasi	40
4.1.1	Hasil Matrix <i>Perplexity</i>	40
4.1.2.	Hasil <i>Generate Pantun</i>	41
4.2.	<i>Deployment</i>	44
4.2.1.	Olahan lanjutan dari hasil <i>generate pantun</i>	44
4.2.2.	Tampilan <i>Deployment</i>	46
BAB 5 PENUTUP		50

5.1. Kesimpulan.....	50
5.2. Saran.....	50
DAFTAR REFERENSI	52
LAMPIRAN.....	56

DAFTAR GAMBAR

Gambar 2.1 Arsitektur <i>Transformers</i> (Vaswani, dkk., 2017).....	13
Gambar 2.2 <i>Scaled Dot-Product Attention</i> (Vaswani, dkk., 2017)	14
Gambar 2.3 <i>Multi-Head Attention</i> (Vaswani, dkk., 2017).....	15
Gambar 3.1 Tahapan Penelitian	20
Gambar 3.2 17 Folder Jenis Pantun	26
Gambar 3.3 Hasil Proses <i>Line Merging</i> dan <i>Escape Sequence</i>	31
Gambar 3.4 Hasil Penyatuan ke-17 Folder Jenis Pantun	33
Gambar 3.5 Pembagian umum parameter untuk setiap ukuran GPT-2 (Radford, dkk., 2019)	35
Gambar 4.1 Tampilan awal Generator Pantun.....	46
Gambar 4.2 Tampilan input “contoh” Generator Pantun.....	47
Gambar 4.3 Tampilan meminta “isi” Generator Pantun	48
Gambar 4.4 Tampilan meminta “sampiran” Generator Pantun	49

DAFTAR TABEL

Tabel 1.1 <i>State of The Art</i>	3
Tabel 3.1 Kesulitan dalam pembuatan pantun (Pebrianto, 2017)	22
Tabel 3.2 Distribusi Pantun.....	26
Tabel 3.3 Tabel Hasil Normalisasi Kata	28
Tabel 3.4 Tabel Hasil <i>Punctuation Removal</i>	28
Tabel 3.5 Tabel Hasil <i>Line Merging</i>	30
Tabel 3.6 Tabel Hasil Penambahan <i>Escape Sequence</i>	30
Tabel 3.7 Folder Jenis Pantun dan <i>File</i> di dalamnya	32
Tabel 3.8 Tabel Hasil Penambahan <BOS> dan <EOS>	33
Tabel 3.9 Pembagian <i>Training Data</i> dan <i>Test Data</i>	34
Tabel 3.10 Hasil Model <i>Pre-Trained</i> Indonesian-nlp (Sahid, dkk., 2022)	35
Tabel 3.11 Hasil <i>Text Generation</i> dari Model Indonesian-nlp	36
Tabel 3.12 Modifikasi <i>Hyperparameter</i>	38
Tabel 3.13 Hasil <i>Modelling Epoch 3</i>	39
Tabel 4.1 Evaluasi Hasil <i>Perplexity</i> pada epoch 2, 3, 4, 5, dan 10.....	40
Tabel 4.2 Parameter untuk <i>Generate</i> Pantun	41
Tabel 4.3 Hasil <i>Generate</i> Pantun	41
Tabel 4.4 Hasil <i>Generate</i> Pantun yang sudah diolah.....	42
Tabel 4.5 Pantun ke-10 dari Hasil <i>Generate</i> Pantun.....	44
Tabel 4.6 Contoh <i>Case</i> Kejadian Pantun yang identik	44
Tabel 4.7 Hasil <i>Generate</i> Pantun olahan lanjutan	45
Tabel 4.8 Pemahaman <i>Indexing</i> dari Olahan	45