

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Pantun adalah salah satu jenis puisi rakyat Indonesia yang terdiri dari empat baris, di mana setiap baris maksimal terdiri atas 12 suku kata. Pantun memiliki rima akhir yang sama pada setiap barisnya. (Setyadiharja, 2020). Struktur barisnya adalah baris pertama dan kedua disebut dengan “sampiran” dan baris ketiga dan keempat disebut dengan “isi”. (Setyadiharja, 2020). Pantun biasanya digunakan sebagai bentuk hiburan dalam acara-acara tertentu seperti disisipi dalam pernikahan, pertunjukan seni, atau acara keagamaan. (Ridwan, 2022). Selain itu, pantun juga digunakan sebagai sarana penyampaian pesan atau nilai-nilai kehidupan seperti nasihat, perumpamaan, atau bahkan sindiran. Hal itu menunjukkan bahwa pantun bisa dilakukan dalam keadaan bagaimanapun dengan catatan pantun yang diberikan harus disesuaikan dengan kebutuhan saat itu. (Rustiawati, 2016).

Pada proses pembuatan pantun berbahasa Indonesia terdapat beberapa masalah yang harus dihadapi seperti keterbatasan pengetahuan kata, variasi yang terbatas, serta kesulitan dalam menjaga konsistensi pola dan sajak. (Pebrianto, 2017). Kesulitan tersebut terbentuk dikarenakan pantun sendiri ialah bentuk puisi tradisional yang memiliki aturan sajak dan rima tertentu. Semua permasalahan ini dapat menghambat kreativitas dalam proses pembuatan pantun berbahasa Indonesia.

Dalam perkembangan teknologi yang semakin maju, kini telah hadir GPT (*Generative Pre-trained Transformers*), suatu arsitektur model bahasa alami berbasis kecerdasan buatan atau *Artificial Intelligence* yang mampu menghasilkan teks yang menyerupai hasil penulisan manusia dengan tingkat akurasi yang tinggi. (Brown, dkk., 2020). GPT dapat digunakan untuk berbagai tugas *Natural Language Processing* seperti menjawab pertanyaan, menyimpulkan teks, dan sebagainya. (Shree, 2020). Dengan menggabungkan teknologi GPT dan pantun berbahasa Indonesia, maka dapat dikembangkan sebuah aplikasi yang mengimplementasikan kemampuan GPT sebagai pantun generator. Dengan adanya aplikasi GPT pantun generator ini, diharapkan dapat membantu menuntaskan permasalahan dalam pembuatan pantun dan melestarikan kebudayaan Indonesia, khususnya dalam bidang sastra dan puisi tradisional. Selain itu, aplikasi ini juga dapat dimanfaatkan sebagai

media hiburan, edukasi, dan bahkan dalam industri kreatif seperti iklan, konten digital, dan sebagainya.

Penelitian sebelumnya telah dilakukan terkait pembuatan pantun generator. Penelitian tersebut dilakukan oleh Helena Nurramdhani Irmada, Ria Astriratma, Nurul Chamidah, dan Mayanda Mega Santoni pada tahun 2021. Penelitian tersebut mengusulkan “Pembuatan Sampiran Pantun Otomatis Berbasis *Pattern-Matching*”. Hasil dari penelitian tersebut menunjukkan bahwa metode *pattern-matching* dapat digunakan untuk membuat teks sampiran pantun secara otomatis sesuai dengan kaidah pantun, baik secara jumlah larik maupun rimanya. Hal ini sejalan dengan hasil evaluasi *naturalness* yang baik dari pengguna dalam aspek *readability*, *clarity*, dan *general appropriateness* yang masing-masing mencapai nilai evaluasi sebesar 95%, 93%, dan 97,5%.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dituliskan topik ini mengangkat beberapa rumusan masalah, diantaranya:

1. Bagaimana cara mengimplementasikan model *Generative Pre-trained Transformers* (GPT) sebagai pantun generator?
2. Bagaimana cara melatih model GPT agar dapat menghasilkan pantun-pantun berbahasa Indonesia?
3. Seberapa baik kinerja dari GPT pantun generator dalam menghasilkan pantun-pantun berbahasa Indonesia?

1.3. Tujuan Penelitian

Beberapa tujuan yang ingin dicapai setelah pengerjaan tugas akhir diantaranya:

1. Mengimplementasikan model *Generative Pre-trained Transformers* (GPT) sebagai pantun generator.
2. Melatih model GPT agar dapat menghasilkan pantun-pantun berbahasa Indonesia.

3. Menghasilkan sebuah aplikasi GPT pantun generator yang dapat memproses bahasa alami dan menghasilkan pantun-pantun berbahasa Indonesia secara otomatis.

1.4. Ruang Lingkup

Tugas akhir ini memiliki ruang lingkup sebagai berikut:

1. Teks yang digunakan adalah teks berbahasa Indonesia.
2. Data yang digunakan akan terdiri dari pantun-pantun berbahasa Indonesia dari berbagai jenis pantun yang dikumpulkan dan diolah sesuai dengan format yang dibutuhkan oleh GPT.
3. Penelitian akan berfokus pada implementasi teknologi kecerdasan buatan menggunakan GPT sebagai pantun generator.
4. Evaluasi model menggunakan metrik *perplexity*.
5. Model GPT pantun generator yang dikembangkan dalam tugas akhir ini akan memiliki fitur untuk memproses masukan berupa kalimat yang diinginkan oleh pengguna dan menghasilkan keluaran berupa pantun-pantun berbahasa Indonesia.
6. Menggunakan bahasa pemrograman Python dan beberapa *library* yang dibutuhkan untuk mengintegrasikan GPT ke dalam aplikasi tersebut.

1.5. State of The Art

Tabel 1.1 *State of The Art*

Judul Jurnal	Pembahasan
<p>Pembuat Sampiran Pantun Otomatis berbasis <i>Pattern-matching</i></p> <p>Peneliti Helena Nurramdhani Irmanda, Ria Astriratma, Nurul Chamidah, Mayanda Mega Santoni</p> <p>Lokasi</p>	<p><u>Hasil Penelitian:</u> Jurnal ini berisi tentang Metode <i>pattern-matching</i> untuk membuat teks sampiran pantun secara otomatis. Hasil evaluasi akhir memiliki nilai <i>naturalness</i> yang baik. Hal ini dibuktikan dengan hasil dari ketiga aspek evaluasi yaitu <i>readability</i>, <i>clarity</i>, dan <i>general appropriateness</i> yang</p>

<p>Universitas Pembangunan Nasional Veteran Jakarta</p> <p>Tahun 2021</p> <p>Publisher Jurnal SISFOKOM (Sistem Informasi dan Komputer) / Lembaga Penelitian dan Pengabdian Masyarakat (LPPM)</p>	<p>cukup tinggi masing-masing sebesar 95%, 93% dan 97,5%.</p> <p><u>Alasan Menjadi Tinjauan Penelitian:</u> Jurnal ini memiliki relevansi dengan tujuan yang sama yaitu membuat Pantun Otomatis tetapi memakai metode dan <i>tools</i> yang berbeda</p>
<p><i>Improving Language Understanding by Generative Pre-Training</i></p> <p>Peneliti Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever</p> <p>Tahun 2018</p> <p>Publisher OpenAI</p>	<p><u>Hasil Penelitian:</u> Jurnal ini menunjukkan efektivitas <i>pre-training</i> generatif diikuti dengan <i>fine-tuning</i> diskriminatif dalam meningkatkan pemahaman bahasa alami. Dengan melakukan <i>pre-training</i> pada model bahasa menggunakan korpus teks tak berlabel atau <i>unsupervised learning</i> yang beragam, dan kemudian melakukan <i>fine-tuning</i> pada tugas-tugas spesifik, Hasil penelitian menegaskan efektivitas <i>pre-training</i> generatif dalam meningkatkan pemahaman bahasa, dan menjadikannya sebagai dasar yang berharga untuk pengembangan model pemahaman bahasa yang lebih lanjut.</p> <p><u>Alasan Menjadi Tinjauan Penelitian:</u> Jurnal ini menyajikan hasil penelitian yang relevan dengan penelitian yang akan dilakukan, yaitu peningkatan</p>

	<p>pemahaman bahasa menggunakan pendekatan <i>generative pre-training</i>. Jurnal ini memberikan bukti bahwa pendekatan tersebut berhasil meningkatkan kinerja pemahaman bahasa secara signifikan.</p>
<p><i>Language Models are Unsupervised Multitask Learners</i></p> <p>Peneliti Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever</p> <p>Tahun 2019</p> <p>Publisher OpenAI</p>	<p><u>Hasil Penelitian:</u></p> <p>Jurnal ini menunjukkan bahwa model dalam penelitian ini, GPT-2 dengan 1,5 miliar parameter <i>Transformers</i>, mencapai hasil terbaik pada 7 dari 8 dataset pemodelan bahasa yang diuji. Temuan ini memberikan arah yang menjanjikan dalam pengembangan sistem pemrosesan bahasa menggunakan GPT-2.</p> <p><u>Alasan Menjadi Tinjauan Penelitian:</u></p> <p>Jurnal ini memberikan kontribusi penting dalam memahami kemampuan model bahasa dalam konteks pemrosesan bahasa alami menggunakan GPT-2. Jurnal ini mengungkapkan bahwa model GPT-2 menunjukkan performa yang baik. Hal ini berkaitan dengan penelitian pembuatan pantun, karena penelitian ini memberikan wawasan tentang bagaimana model bahasa dapat digunakan untuk menghasilkan teks yang berkualitas dan konsisten dengan pola bahasa yang diinginkan.</p>

<p><i>Language Models are Few-Shot Learners</i></p> <p>Peneliti</p> <p>Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, Dario Amodei</p> <p>Tahun</p> <p>2020</p> <p>Publisher</p> <p>OpenAI</p>	<p><u>Hasil Penelitian:</u></p> <p>Jurnal ini menunjukkan bahwa model GPT-3, yang merupakan model Bahasa dengan 175 miliar parameter, mampu mencapai performa yang baik dengan hanya menggunakan sedikit data pelatihan atau <i>few-shot learning</i>. GPT-3 berhasil mencapai performa yang kompetitif dengan pendekatan <i>fine-tuning</i> terkini, bahkan tanpa melakukan pembaruan gradien atau <i>fine-tuning</i> khusus pada tugas tertentu. Model ini mampu mengatasi berbagai tugas NLP, termasuk terjemahan, pertanyaan-jawaban, serta tugas yang melibatkan penalaran dan adaptasi domain secara dinamis. Hasil penelitian ini juga menunjukkan bahwa GPT-3 mampu menghasilkan sampel artikel berita yang sulit dibedakan oleh evaluator manusia. Temuan ini memiliki dampak yang luas dalam masyarakat dan perlu dipertimbangkan secara serius.</p> <p><u>Alasan Menjadi Tinjauan Penelitian:</u></p> <p>Jurnal ini mengungkap potensi yang besar dari model bahasa dalam melakukan tugas-tugas <i>few-shot</i> dengan kinerja yang kuat. Dalam penelitian ini, model GPT-3 dengan skala yang besar sehingga memberikan wawasan dalam pengembangan model bahasa yang</p>
--	---

	<p>mampu belajar dari sedikit data pelatihan. Penelitian yang sedang dilakukan dapat mengambil inspirasi dan pemahaman lebih lanjut dalam membangun model bahasa yang efektif.</p>
<p><i>GPoeT-2: A GPT-2 Based Poem Generator</i></p> <p>Peneliti Kai-Ling Lo, Rami Ariss, Philipp Kurz</p> <p>Lokasi Carnegie Mellon University</p> <p>Tahun 2022</p> <p>Publisher arXiv (<i>Computation and Language Journal</i>)</p>	<p><u>Hasil Penelitian:</u></p> <p>Jurnal ini berisi tentang pembuatan sebuah sistem bernama GPoeT-2 yang dapat menghasilkan puisi menggunakan model bahasa GPT-2. Sistem ini fokus pada pembuatan <i>limericks</i>, yaitu puisi terstruktur yang sering kali mengandung humor. GPoeT-2 menggunakan teknik <i>fine-tuning</i> untuk menghasilkan <i>limericks</i> yang mengikuti pola berima AABBA. Evaluasi dilakukan dengan memperhatikan kebenaran tata bahasa, keberagaman kata, dan kesinambungan tema. Tujuan utama dari penelitian ini adalah untuk memicu kreativitas manusia dalam menghasilkan puisi.</p> <p><u>Alasan Menjadi Tinjauan Penelitian</u></p> <p>Jurnal ini memberikan wawasan dan inspirasi tentang penerapan model bahasa dalam menciptakan karya sastra, khususnya dalam penghasilan sajak dalam puisi.</p>

1.6. Sistematika Penulisan

Agar dapat lebih jelas dalam memahami laporan tugas akhir ini, maka materi-materi yang tertera pada laporan dikelompokkan menjadi beberapa bab. Sistematika penulisan Tugas Akhir adalah sebagai berikut:

- Bab 1** **Pendahuluan**
Bab pendahuluan berisi tentang latar belakang, rumusan masalah, tujuan penelitian, ruang lingkup, *state of the art*, dan sistematika penulisan.
- Bab 2** **Landasan Teori**
Bab landasan teori berisi tentang landasan teori penelitian yang terdiri dari teori dasar *machine learning*, *deep learning*, *preprocessing data*, *transformers*, *generative pre-trained transformers*, *Pytorch* dan *streamlit*.
- Bab 3** **Metodologi Penelitian**
Bab metodologi penelitian berisi tentang analisis dan perancangan mengenai dataset dan model *generative pre-trained transformers* yang akan digunakan.
- Bab 4** **Hasil dan Pembahasan**
Bab hasil dan pembahasan berisi tentang *fine-tuning* model *generative pre-trained transformers* menggunakan dataset pantun berbahasa Indonesia, evaluasi dan pengujian model.
- Bab 5** **Penutup**
Bab penutup berisi kesimpulan dan saran dari pembahasan hasil dan pengujian pada implementasi *generative pre-trained transformers* sebagai pantun generator.