

PENERAPAN ALGORITMA TRANSFORMER DALAM APLIKASI PARAFRASE TEKS OTOMATIS

Robet, Kelvin Leonardi Kohsasih, Jenime Darwin✉

Program Studi Teknik Informatika, STMIK TIME, Medan, Indonesia

Email: jenimelin20607@gmail.com

ABSTRACT

The development of Natural Language Processing (NLP) technology has enabled the creation of automated text manipulation applications, one of which is text paraphrasing. This study aims to implement a Transformer architecture with a focus on Indonesian text for automatic text paraphrasing applications. The model used is a pre-trained Text-to-Text Transfer Transformer (T5), which is fine-tuned using an Indonesian text corpus called the Indo-T5 model. During the training process, the model is trained to understand language structure and context in order to generate paraphrases that are not only grammatically correct but also semantically preserved. Evaluation was conducted using BLEU and ROUGE metrics to measure the similarity between the generated paraphrased texts and manual references. The evaluation results show that the model is capable of producing coherent, relevant paraphrased texts with a good level of lexical variation with a BLEU score of 50.1, and ROUGE-L of 61.7. Thus, this study demonstrates that Transformer-based models can be effectively applied to the task of text paraphrasing in Indonesian.

Keyword: NLP, Indo-T5, Text Paraphrasing, BLEU, ROUGE.

ABSTRAK

Perkembangan teknologi Natural Language Processing (NLP) telah memungkinkan pengembangan aplikasi manipulasi teks secara otomatis, salah satunya dalam bentuk parafrase teks. Penelitian ini bertujuan untuk mengimplementasikan arsitektur Transformer dengan fokus pada teks Bahasa Indonesia untuk aplikasi parafrase teks otomatis. Model yang digunakan adalah pre-trained Text-to-Text Transfer Transformer (T5) yang dilakukan fine-tuning menggunakan korpus teks Bahasa Indonesia yang dinamakan dengan model Indo-T5. Dalam proses pelatihan, model dilatih untuk memahami struktur Bahasa serta konteks, agar mampu menghasilkan parafrase yang tidak hanya gramatikal namun juga semantiknya tetap terjaga. Evaluasi dilakukan menggunakan metrik BLEU, ROUGE untuk mengukur kesamaan antara teks hasil parafrase dengan referensi manual. Hasil evaluasi menunjukkan bahwa model mampu menghasilkan teks parafrase yang koheren, relevan, dan memiliki tingkat variasi leksikal yang baik dengan nilai BLEU sebesar 50.1, dan ROUGE-L sebesar 61.7. Dengan demikian, penelitian ini membuktikan bahwa pemanfaatan model Transformer dapat diaplikasikan secara efektif untuk tugas parafrase teks dalam Bahasa Indonesia.

Kata Kunci: NLP, Indo-T5, Parafrase Teks, BLEU, ROUGE.

PENDAHULUAN

Kemajuan teknologi informasi dan kecerdasan buatan (AI) dalam dekade terakhir telah membawa perubahan signifikan dalam berbagai aspek kehidupan, termasuk dalam pengolahan Bahasa alami atau *Natural Language Processing (NLP)* (Purnama & Widya Utami, 2023). NLP memungkinkan mesin memahami, menganalisis, serta menghasilkan bahasa manusia secara otomatis. Salah satu penerapan penting dari NLP adalah dalam bidang parafrase teks, yaitu proses mengubah suatu kalimat atau paragraf menjadi bentuk lain dengan makna yang tetap sama namun menggunakan struktur dan kosa kata yang berbeda untuk menghindari plagiarisme dan memperkaya kosa kata yang digunakan (Wahyudi et al., 2023)(Nisa et al., 2022).

Perkembangan arsitektur *Transformer*, yang diperkenalkan oleh (Vaswani et al., 2017), telah merevolusi pendekatan dalam berbagai tugas NLP. Dengan mekanisme *attention* yang inovatif, model berbasis *Transformer* seperti *BERT (Bidirectional Encoder Representation from Transformers)*, *GPT (Generative Pre-Trained Transformer)*, dan *T5 (Text-to-Text Transfer Transformer)* mampu mencapai kinerja tinggi dalam berbagai aplikasi, mulai dari penerjemahan mesin, penjawaban soal, hingga parafrase teks (Bahani et al., 2023). Model-model ini tidak hanya unggul dalam akurasi, tetapi juga memiliki kemampuan generalisasi yang baik terhadap berbagai jenis bahasa dan konteks.

Namun demikian, penelitian tentang penerapan teknologi tersebut di Indonesia masih relatif terbatas,

terutama untuk Bahasa Indonesia. Hal ini disebabkan oleh beberapa faktor, antara lain kurangnya korpus data teks berkualitas dalam skala besar, kompleksitas morfologi dan sintaksis Bahasa Indonesia (Koto et al., 2020), serta minimnya penelitian dan pengembangan aplikasi berbasis *Transformer* yang spesifik untuk Bahasa Nasional. Padahal, kebutuhan akan sistem parafrase otomatis dalam Bahasa Indonesia sangat besar, terutama dalam bidang pendidikan, jurnalistik, penulisan kreatif, serta pengembangan konten digital. Dalam dunia pendidikan, misalnya, sistem parafrase otomatis dapat digunakan sebagai alat bantu menulis untuk menghindari plagiarisme, memberikan alternatif formulasi jawaban, atau meningkatkan variasi ekspresi bahasa siswa (Januarahman & Romadhony, 2023). Di sektor media dan jurnalistik, aplikasi semacam ini dapat membantu redaktur dalam membuat versi berita yang lebih variatif tanpa mengubah pesan inti. Selain itu, dalam konteks layanan publik dan pengabdian masyarakat, sistem ini dapat mendukung penyerdehanaan informasi bagi kelompok masyarakat tertentu, seperti anak-anak atau penyandang disabilitas (Kartika et al., 2023).

Oleh karena itu, penelitian ini bertujuan untuk mengembangkan sebuah aplikasi parafrase teks otomatis berbasis *Transformer* yang dikhususkan untuk Bahasa Indonesia. Aplikasi ini diharapkan dapat menjawab kebutuhan akan alat yang mampu menghasilkan parafrase yang akurat, kontekstual, dan sesuai dengan aturan gramatikal Bahasa Indonesia.

TINJAUAN PUSTAKA

Parafrase Teks

Parafrase teks merupakan salah satu teknik dalam pengolahan bahasa alami yang bertujuan untuk mengungkapkan kembali sebuah kalimat atau paragraf dengan menggunakan kata-kata, frasa, serta struktur tata bahasa yang berbeda, tanpa mengubah makna inti dari teks asli (Lubis et al., 2023). Tujuan utama dari parafrase adalah menjaga keakuratan informasi sembari memberikan variasi ekspresi, sehingga hasil tulisan menjadi lebih mudah dipahami, atau sesuai dengan konteks baru.

Proses parafrase tidak hanya melibatkan penggantian sinonim saja, tetapi juga mencakup perubahan struktur sintaksis, pengaturan ulang urutan kalimat, penggunaan bentuk aktif-pasif, serta penyerdehanaan frasa kompleks tanpa menghilangkan makna semula (Syahnaz & Fithriani, 2023).

Deep Learning

Deep Learning (DL) merupakan cabang dari *machine learning* yang memanfaatkan algoritma

berbasis *neural networks* untuk merepresentasikan dan menganalisis data yang kompleks. Pendekatan ini bertujuan untuk meniru mekanisme pemrosesan informasi pada otak manusia, yang dikenal dengan istilah *artificial neural networks* (Schmidhuber, 2015).

Transformer

Transformer pertama kali dikenalkan oleh Vaswani dan rekan-rekannya pada tahun 2017 dalam makalah berjudul *Attention is All You Need*. Model ini dikembangkan untuk menyelesaikan berbagai permasalahan dalam pemrosesan bahasa alami (*Natural Language Processing* atau NLP), dan sejak saat itu menjadi fondasi bagi beragam model mutakhir seperti *GPT* dan *BERT* (Vaswani et al., 2017).

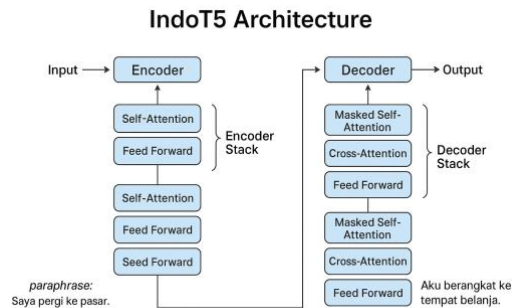
Natural Language Processing (NLP)

Natural Language Processing (NLP) adalah cabang dari kecerdasan buatan (AI) yang berfokus pada interaksi antara komputer dan bahasa manusia. Tujuan utama NLP adalah memungkinkan komputer untuk memahami, menganalisis, dan menghasilkan bahasa yang digunakan oleh manusia secara alami (Hikmah et al., 2023).

Text-to-Text Transfer Transformer (T5)

Text-to-Text Transfer Transformer (*T5*) adalah arsitektur *deep learning* yang mengonversi setiap tugas pemrosesan Bahasa alami ke dalam format *text-to-text*, di mana baik input maupun output direpresentasikan sebagai teks. Pendekatan ini memungkinkan model untuk menangani berbagai tugas NLP seperti terjemahan, klasifikasi, ekstraksi informasi, dan paraphrase hanya dengan pelatihan tunggal pada satu arsitektur yang seragam. Keunggulan *T5* terletak pada kemampuannya untuk memanfaatkan *transfer learning* dalam skala besar dengan pra-pelatihan menggunakan *Corpus C4 (Colossal Clean Crawled Corpus)*, dan kemudian *fine-tuning* pada tugas spesifik. Dengan menyatukan berbagai tugas NLP dalam satu kerangka kerja, *T5* menyerdehanakan arsitektur dan meningkatkan generalisasi lintas tugas, menjadikannya salah satu model paling fleksibel dalam bidang NLP (Raffel et al., 2020), sedangkan *Indo-T5* adalah versi adaptasi dari arsitektur *T5* yang dilatih secara khusus untuk bahasa Indonesia. Seperti *T5* asli, *Indo-T5* memformulasikan semua tugas pemrosesan Bahasa alami dalam format *text-to-text*, di mana input dan output keduanya berupa teks. Model ini dilatih menggunakan data berbahasa Indonesia dalam skala besar agar mampu memahami dan menghasilkan teks yang lebih akurat dan kontekstual sesuai dengan

struktur dan karakteristik linguistic Bahasa Indonesia (Wongso et al., 2023).



Gambar 1. Arsitektur Model Indo-T5

Evaluasi Metrik BLEU dan ROUGE

Evaluasi metrik *BLEU* dan *ROUGE* adalah metode untuk mengukur kualitas output teks yang dihasilkan oleh Bahasa (seperti terjemahan mesin, ringkasan otomatis, atau generasi teks) dengan membandingkannya terhadap referensi manusia. *BLUE* (*Bilingual Evaluation Understudy*) membandingkan n-gram urutan kata antara output model dan referensi manusia (Chris Callison-Burch Miles Osborne Philipp Koehn, 2006), sedangkan *ROUGE* (*Recall Oriented Understudy for Gisting Evaluation*) mengevaluasi kualitas parafrase otomatis dengan mengukur seberapa banyak n-gram atau unit kata dalam referensi yang tercakup oleh output model (Ganesan, 2018).

PyCharm

PyCharm merupakan sebuah *Integrated Development Environment* (IDE) yang dibuat secara khusus untuk mendukung pengembangan menggunakan bahasa pemrograman Python. Dikembangkan oleh JetBrains (Hu et al., 2018), *PyCharm* tersedia dalam dua varian, yaitu *Community Edition* yang bersifat gratis dan *Professional Edition* dengan fitur lebih lengkap. IDE ini menawarkan berbagai fitur yang dirancang untuk membantu pengembang dalam menulis kode secara lebih efisien dan meningkatkan kualitas pengembangan perangkat lunak.

Flask

Flask merupakan *micro web framework* berbasis Python yang dirancang untuk mempermudah pengembangan aplikasi web secara efisien tanpa mengorbankan fleksibilitas dan kapabilitasnya (Walingkas & Saian, 2023). *Framework* ini mendukung integrasi dengan berbagai ekstensi pihak ketiga yang dapat memperluas fitur, seperti sistem autentikasi dan pembuatan *REST API*. Beberapa keunggulan *Flask* antara lain kemudahan penggunaan,

struktur yang fleksibel dan modular, performa yang baik, serta kompatibilitas tinggi dengan berbagai pustaka *Python*.

METODE PENELITIAN

Dataset dan Variasi Data

Dataset yang digunakan dalam pelatihan model ini terdiri dari lebih dari 20.000 pasangan kalimat Bahasa Indonesia hasil parafrase, yang berasal dari berbagai sumber seperti artikel berita, dialog publik, dan dataset parafrase terbuka dari *HuggingFace* (cahya/gpt2-large-indonesian-522M).

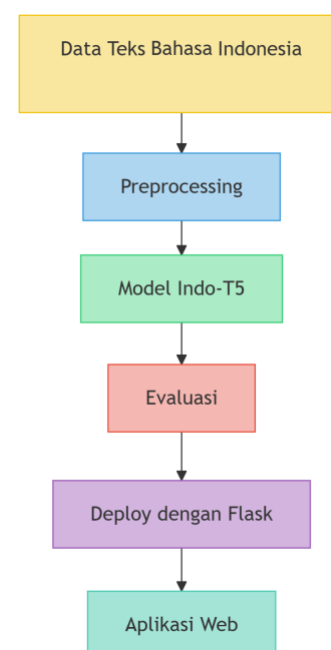
Dataset mencakup variasi panjang kalimat, struktur sintaksis, dan gaya bahasa yang beragam, mencerminkan berbagai domain penggunaan seperti edukasi, hukum, dan hiburan.

Prosedur Validasi dan Kurasi Data

Sebelum digunakan dalam pelatihan, data melalui proses kurasi manual dan semi-otomatis. Data dibersihkan dari entri yang duplikat atau terlalu mirip (plagiarisme langsung). Sebanyak 20% data diperiksa secara manual oleh evaluator untuk memastikan pasangan kalimat merupakan parafrase sah dengan tingkat kesetaraan makna yang tinggi.

Penelitian ini menggunakan metode kuantitatif dengan metode eksperimen untuk mengembangkan dan mengevaluasi sistem parafrase otomatis berbasis model *Indo-T5*. Eksperimen ini memanfaatkan model prelatih *IndoT5-base paraphrase* yang tersedia melalui repositori *HuggingFace*.

Tahapan dalam penelitian ditunjukkan pada gambar 2 di bawah ini.



Gambar 2. Metode Penelitian

1. Data
Dataset yang digunakan terdiri dari pasangan kalimat asli dan hasil parafrase dalam Bahasa Indonesia, yang dikumpulkan dari berbagai sumber seperti korpus publik, artikel berita, dan dataset parafrase Indonesia
2. *Preprocessing* data adalah langkah penting dalam pengembangan model parafrase. Tujuan dari *preprocessing* data adalah untuk membersihkan teks dari karakter yang tidak diinginkan. Berikut adalah langkah yang dilakukan dalam *preprocessing* data:
 - a. Membersihkan teks dari karakter yang tidak diinginkan seperti: tanda baca, simbol, dan karakter spesial yang dapat mengganggu kinerja model.
 - b. Tokenisasi: tokenisasi adalah proses memecah teks menjadi token-token yang lebih kecil. Token-token ini dapat berupa kata, frasa, atau kalimat.
 - c. Normalisasi: Normalisasi adalah proses mengubah teks menjadi bentuk yang lebih sederhana. Contohnya, mengubah kata "kamu" menjadi "anda".
 - d. *Padding*: *padding* adalah proses menambahkan karakter-karakter kosong ke teks agar panjang teks menjadi sama. Hal ini diperlukan karena model memerlukan input dengan panjang yang sama.
3. Setelah *preprocessing* selesai dilakukan, maka tahapan selanjutnya adalah melakukan pelatihan model dengan mengimplementasikan *AutoModelForSeq2SeqLM* dari pustaka *Hugging Face* dan *fine-tuning* parameter dilakukan terhadap dataset parafrase dengan menggunakan konfigurasi sebagai berikut:
 $max_length = 512$,
 $do_sample = True$,
 $top_k = 200$,
 $top_p = 0.95$,
 $early_stopping = True$.
4. Evaluasi
Evaluasi dilakukan dengan menggunakan dua metrik utama, yaitu : *BLUE* dan *ROUGE*.
5. Hasil dari pelatihan model dan evaluasi tersebut, kemudian diintegrasikan ke aplikasi berbasis web menggunakan *Flask*.

Adapun langkah-langkah dalam proses perancangan aplikasi adalah sebagai berikut :

- a. Arsitektur Sistem:
 - 1) Aplikasi web terdiri dari dua bagian utama: *frontend* (antarmuka pengguna) dan *backend* (server aplikasi).

- 2) *Frontend* menerima *input* dari pengguna dan menampilkan hasil parafrase.
- 3) *Backend* memproses *input* menggunakan model *IndoT5-base-parafrase* dan mengirim hasilnya kembali ke *frontend*.

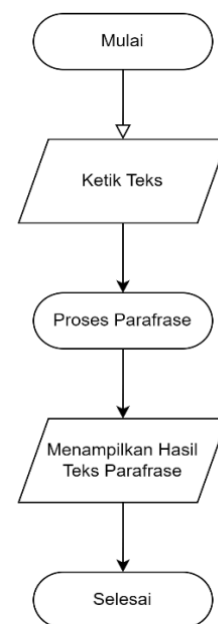
b. Komponen Utama Sistem:

- 1) *Frontend*: dibuat dengan *HTML*, *CSS*, dan *JavaScript*.
- 2) *Backend*: dikembangkan dengan menggunakan *Flask*. Mengintegrasikan model *IndoT5-base-parafrase* untuk memproses teks yang diterima, dan kemudian model melakukan parafrase.

c. Desain Alur Kerja

1) *Flowchart*

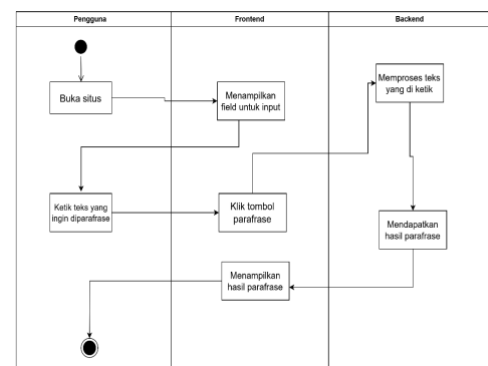
Flowchart ini menggambarkan proses dari *input* pengguna hingga keluarnya hasil parafrase:



Gambar 3. Alur Parafrase

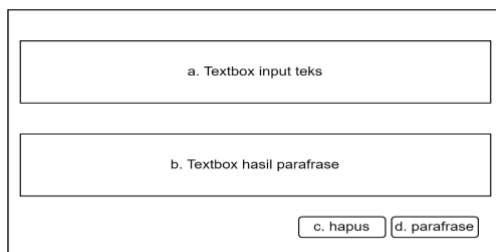
2) *Activity Diagram*

Activity diagram ini menggambarkan interaksi antara pengguna, *frontend*, dan *backend* :



Gambar 4. Activity Diagram

- Pengguna memasukkan teks yang akan diparafrase melalui *form* di halaman web.
 - Teks dikirim ke *backend* melalui *API*.
 - Backend* melakukan tokenisasi dan pemrosesan teks menggunakan model *IndoT5-base-paraphrase* dan menghasilkan hasil parafrase.
 - Hasil parafrase dikirim kembali ke *frontend* dan ditampilkan kepada pengguna.
- d. Desain Tampilan Antarmuka
- Tampilan utama adalah tampilan yang pertama kali muncul ketika situs diakses, pada tampilan utama terdapat beberapa tampilan seperti terlihat pada Gambar dibawah ini.



Gambar 5. Tampilan Antarmuka

Keterangan:

- 1) *Textbox input* teks yang disediakan untuk mengisi teks yang ingin diparafrase.
- 2) *Textbox* hasil parafrase untuk menampilkan hasil parafrase.
- 3) Tombol Hapus digunakan untuk menghapus seluruh teks yang telah diketik.
- 4) Tombol Parafrase berfungsi untuk konfirmasi parafrase dari *textbox input* teks.

HASIL DAN PEMBAHASAN

Detail Konfigurasi Pelatihan

Model dilatih menggunakan *batch size* sebesar 16, learning rate $5e-5$, dan *optimizer AdamW*. Proses pelatihan dilakukan selama 5 *epoch* dengan pembagian data 80% untuk data latih dan 20% untuk data validasi. Proses *fine-tuning* dijalankan menggunakan *GPU NVIDIA RTX 3050* pada lingkungan *Python 3.10* dengan Pustaka *HuggingFace Transformers*.

Evaluasi Manual dan Analisis Kesalahan

Selain evaluasi otomatis menggunakan *BLEU* dan *ROUGE*, dilakukan evaluasi subjektif oleh 10 evaluator yang menilai kualitas parafrase berdasarkan keterbacaan, kealamian bahasa, dan kesesuaian makna. Evaluator memberikan skor pada skala 1–5 untuk setiap dimensi tersebut. Hasil menunjukkan bahwa 84% kalimat parafrase dinilai setara secara semantik dan 78% mudah dibaca.

Analisis kesalahan dilakukan terhadap sampel kalimat dengan skor *BLEU* < 10. Ditemukan bahwa

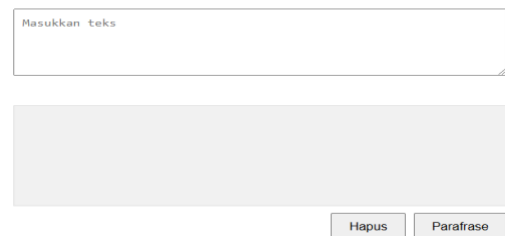
kegagalan model sering terjadi pada kalimat idiomatik, struktur bersarang, dan kalimat majemuk bertingkat. Contoh kasus: kalimat “*Ia menjilat ludah sendiri*” diubah menjadi “*Dia minum air yang sudah dibuang*”, yang mengubah makna idiomatic menjadi literal.

Perbandingan Model

Sebagai tambahan, model *IndoT5* dibandingkan dengan *GPT-2* Indonesia pada subset data yang sama. Hasil evaluasi menunjukkan bahwa *IndoT5* menghasilkan skor *BLEU* dan *ROUGE* lebih tinggi rata-rata 10 poin dan memberikan struktur kalimat yang lebih alami.

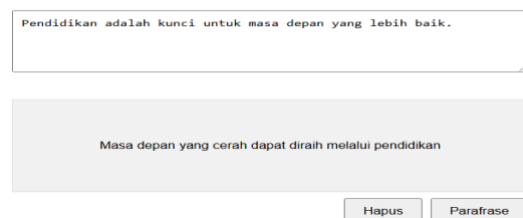
Setelah dilakukan implementasi, maka hasilnya seperti gambar dibawah ini

ALGORITMA TRANSFORMER DALAM PARAFRASE TEKS



Gambar 6. Halaman Utama Aplikasi

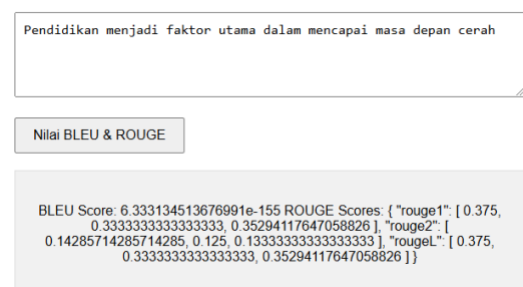
ALGORITMA TRANSFORMER DALAM PARAFRASE TEKS



Gambar 7. Hasil Parafrase

Aplikasi web ini menyediakan antarmuka yang sederhana, mudah digunakan, dan memungkinkan pengguna untuk memasukkan teks, evaluasi model parafrase, dan mendapatkan hasil parafrase dalam waktu singkat.

Evaluasi



Gambar 8. Tampilan Hasil Evaluasi

Hasil evaluasi terhadap model *Indo-T5* dalam menghasilkan parafrase teks menggunakan *BLEU*, dan *ROUGE* ditunjukkan pada Tabel 1 berikut:

Tabel 1. Hasil Nilai Rata-Rata Evaluasi

| Model | BLEU | ROUGE-1 | ROUGE-2 | ROUGE-L |
|----------------|------|---------|---------|---------|
| <i>Indo-T5</i> | 50.1 | 63.4 | 48.2 | 61.7 |

Selain itu, contoh hasil parafrase dari masing-masing model ditampilkan pada Tabel 2 berikut :

Tabel 2. Tabel Hasil Nilai Evaluasi

| Teks Asli | Teks Referensi | Hasil Parafrase | BLEU | ROUGE |
|----------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------|------|-------|
| "Pendidikan adalah kunci untuk masa depan yang lebih baik." | "Pendidikan menjadi faktor utama dalam mencapai masa depan cerah." | "Masa depan yang cerah dapat diraih melalui pendidikan." | 2.3 | 62.1 |
| "Pendidikan membantu seseorang mencapai kesuksesan dalam hidup." | "Dengan pendidikan, seseorang dapat mencapai keberhasilan dalam hidupnya." | "Kesuksesan dalam hidup dapat dicapai dengan pendidikan yang baik." | 0.7 | 60.3 |
| "Teknologi berkembang dengan pesat dalam beberapa dekade terakhir." | "Dalam beberapa tahun terakhir, kemajuan teknologi semakin cepat." | "Dalam beberapa dekade terakhir, teknologi telah berkembang sangat cepat." | 8.9 | 58.5 |
| "Membaca buku dapat meningkatkan wawasan dan keterampilan berpikir." | "Buku adalah sumber pengetahuan yang dapat memperluas wawasan dan meningkatkan kemampuan berpikir." | "Wawasan dan kemampuan berpikir seseorang dapat bertambah dengan membaca buku." | 1.4 | 61.0 |
| "Olahraga secara teratur membantu" | "Rutin berolahraga berperan penting" | "Kesehatan tubuh dan pikiran dapat" | 9.6 | 59.8 |

| | | | | |
|-------------------------------------------|--------------------------------------------|-----------------------------------|--|--|
| ntu menjaga kesehatan tubuh dan pikiran." | dalam menjaga kebugaran fisik dan mental." | dijaga dengan rutin berolahraga." | | |
|-------------------------------------------|--------------------------------------------|-----------------------------------|--|--|

Berdasarkan hasil tersebut, model *Indo-T5* yang telah di-*fine-tuned* menunjukkan performa yang cukup baik dalam menghasilkan teks parafrase yang lebih dekat dengan referensi, berdasarkan metrik *BLEU* dan *ROUGE*.

KESIMPULAN DAN SARAN

Berdasarkan hasil evaluasi yang dilakukan, dapat disimpulkan bahwa:

- Model *IndoT5* yang dituning menunjukkan hasil terbaik dengan nilai *BLEU* sebesar 50.1 dan *ROUGE-L* sebesar 61.7.
- Penggunaan metrik *BLEU* dan *ROUGE* terbukti efektif dalam mengevaluasi kualitas teks parafrase.

- Saran:
- Evaluasi Tambahan menggunakan Metrik lain seperti *METEOR* atau *BERTScore*.
 - Penggunaan Dataset yang lebih luas dan bervariasi.
 - Optimalisasi Model melalui *Fine-Tuning* lebih lanjut.
 - Pengembangan Aplikasi Berbasis Web yang lebih Interaktif.

DAFTAR PUSTAKA

Bahani, M., Ouazizi, A. El, & Maalmi, K. (2023). The effectiveness of T5, GPT-2, and BERT on text-to-image generation task. *Pattern Recognition Letters*, 173, 57–63. <https://doi.org/10.1016/j.patrec.2023.08.001>

Chris Callison-Burch Miles Osborne Philipp Koehn. (2006). Callison-Burch, Ch., Osborne, M., & Koehn, Ph. (2006). Re-evaluating the role of bleu in machine translation. In Proceedings the 11th Conference of the European Chapter of the Association for Computational Linguistics (pp. 249-256). Trento, Italy. *Proceedings the 11th Conference of the European Chapter of the Association for Computational Linguistics*, (pp. 249-256).

Ganesan, K. (2018). *ROUGE 2.0: Updated and Improved Measures for Evaluation of Summarization Tasks*. 1–8.

Hikmah, A., Azmi, F., & Nugrahaeni, R. A. (2023). Implementasi Natural Language Processing Pada Chatbot Untuk Layanan Akademik. *E-Proceeding of Engineering*, 10(1), 371–382.

Januarahman, F., & Romadhony, A. (2023). Paraphrase Generation For Reading

- Comprehension. *Sinkron*, 8(4), 2018–2026.
<https://doi.org/10.33395/sinkron.v8i4.12873>
- Kartika, B. V., Alfredo, M. J., & Kusuma, G. P. (2023). Fine-Tuned IndoBERT based model and data augmentation for indonesian language paraphrase identification. *Revue d'Intelligence Artificielle*, 37(3), 733–743.
<https://doi.org/10.18280/ria.370322>
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP. *COLING 2020 - 28th International Conference on Computational Linguistics, Proceedings of the Conference*, 757–770.
<https://doi.org/10.18653/v1/2020.coling-main.66>
- Lubis, F., Silaban, A. F., Siregar, A. S. A. J., Girsang, A. A., Situmorang, D. N. Br., Purba, G. S. A., Siregar, N. A., & Devi, T. A. (2023). Analisis Pentingnya Parafrase pada Penulisan Artikel Ilmiah Sebagai Upaya Menghindari Plagiarisme. *Jurnal Pendidikan Non Formal*, 1(2), 9.
<https://doi.org/10.47134/jpn.v1i2.150>
- Nisa, K., Syafitri, E., Saragih, S. R. D., Aryni, Y., & Rahmadani, E. (2022). Penggunaan Aplikasi Parafrasa untuk Menurunkan Plagiarisme pada Mahasiswa FKIP Universitas Asahan dalam Menyelesaikan Skripsi. *Jurnal Warta Pengabdian Andalas*, 29(1), 55–60.
<https://doi.org/10.25077/jwa.29.1.55-60.2022>
- Purnama, I. N., & Widya Utami, N. N. (2023). Implementasi Peringkat Dokumen Berbahasa Indonesia Menggunakan Metode Text To Text Transfer Transformer (T5). *Jurnal Teknologi Informasi Dan Komputer*, 9, 4.
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21, 1–67.
- Syahnaz, M., & Fithriani, R. (2023). Utilizing Artificial Intelligence-based Paraphrasing Tool in EFL Writing Class: A Focus on Indonesian University Students' Perceptions. *Scope : Journal of English Language Teaching*, 7(2), 210. <https://doi.org/10.30998/scope.v7i2.14882>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 2017-December(Nips)*, 5999–6009.
- Wahyudi, A., Pamuji, A., & Martuti, R. (2023). Pelatihan Peningkatan Kemampuan Parafrasa Mahasiswa Program Studi Ilmu Keperawatan STIK Bina Husada Palembang Dengan Menggunakan Sinonim Dan Mengubah Struktur Teks. *CARE: Journal Pengabdian Multi Disiplin*, 1(2), 30–36.
<https://doi.org/10.31004/care.v1i2.15402>
- Wongso, W., Joyoadikusumo, A., Buana, B. S., & Suhartono, D. (2023). Many-to-Many Multilingual Translation Model for Languages of Indonesia. *IEEE Access*, 11, 91385–91397.
<https://doi.org/10.1109/ACCESS.2023.3308818>