**PENERAPAN ALGORITMA TRANSFORMER DALAM APLIKASI PARAFRASE TEKS OTOMATIS**

**Jenime Darwin 1 , Octara Pribadi 2 , Hendri 3**

Fakultas Teknik Informatika , STMIK TIME Medan

e-mail: jenime.darwin@gmail.com[1]

**Abstrak**

*Penelitian ini bertujuan untuk merancang dan mengimplementasikan aplikasi parafrase teks otomatis menggunakan algoritma Transformer. Dengan memanfaatkan keunggulan arsitektur Transformer, model mampu menghasilkan parafrase yang akurat dan mempertahankan makna teks asli. Model yang digunakan adalah GPT-2 versi besar yang di-fine-tune pada data teks Bahasa Indonesia. Hasil evaluasi dengan metrik BLEU dan ROUGE menunjukkan bahwa model ini memiliki performa yang baik dalam menghasilkan teks parafrase yang koheren dan relevan.*

*Kata Kunci*: *Transformer*, *GPT-2*, Paraphrase teks, *BLEU, ROUGE, NLP*

***Abstract***

*This research aims to design and implement an automatic text paraphrasing application using the Transformer algorithm. By leveraging the advantages of the Transformer architecture, the model is capable of generating accurate paraphrases while preserving the original meaning. The model used is a large version of GPT-2 that has been fine-tuned on Indonesian text data. Evaluation results using BLEU and ROUGE metrics show that the model performs well in generating coherent and relevant paraphrased text.*

*Keywords: Transformer, GPT-2, Text Paraphrasing, BLEU, ROUGE, NLP*

# Pendahuluan

Kemajuan teknologi dalam bidang *Natural Language Processing* (NLP)telah membuka peluang besar dalam pengembangan aplikasi yang mampu memanipulasi teks secara otomatis. Salah satu inovasi penting dalam NLP adalah arsitektur *Transformer*, yang digunakan secara luas dalam berbagai aplikasi seperti penerjemahan mesin dan parafrase teks. Penelitian ini bertujuan untuk mengembangkan aplikasi parafrase teks otomatis berbasis *Transformer*, dengan fokus pada Bahasa Indonesia

# Landasan Teori

*Transformer* adalah arsitektur *deep learning* yang diperkenalkan oleh Vaswani et al. tahun 2017. Arsitektur ini mengandalkan mekanisme *self-attention* yang memungkinkan model memahami konteks global dari teks input. Dalam konteks parafrase, model seperti GPT-2 dapat di-*fine-tune* untuk menghasilkan versi baru dari teks asli tanpa mengubah makna.

Evaluasi parafrase umumnya dilakukan dengan metrik seperti *BLEU* dan *ROUGE*, yang mengukur kesamaan n-gram dan struktur kalimat antara teks hasil parafrase dan teks referensi.

## Parafrase Teks

Parafrase teks adalah proses mengungkapkan kembali sebuah kalimat atau

paragraf dengan menggunakan kata-kata dan struktur yang berbeda tanpa mengubah makna asli dari teks tersebut

## Deep Learning

*Deep Learning* adalah sub bidang dari *machine learning* yang menggunakan algoritma berbasis jaringan saraf tiruan (*neural networks*) untuk memodelkan dan memahami kompleksitas data. Konsep dasar dari deep learning adalah untuk meniru cara kerja otak manusia dalam mengolah informasi, yang dikenal sebagai *artificial neural networks.*

## Transformer

*Transformer* adalah sebuah arsitektur model *deep learning* yang pertama kali diperkenalkan oleh Vaswani et al. pada tahun 2017 melalui makalah berjudul *Attention is All You Need*. *Transformer* dirancang untuk menangani berbagai tugas pemrosesan bahasa alami (NLP) dan telah menjadi dasar dari banyak model canggih seperti GPT dan BERT.

## 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.

## Generative Pre-trained Transformer (GPT)

*Generative Pre-trained Transformer* (GPT) adalah model bahasa berbasis deep learning yang dikembangkan oleh OpenAI, menggunakan arsitektur *Transformer*.

## Pycharm

PyCharm adalah lingkungan pengembangan terintegrasi (IDE) yang khusus dirancang untuk bahasa pemrograman *Python*.

# Metodologi

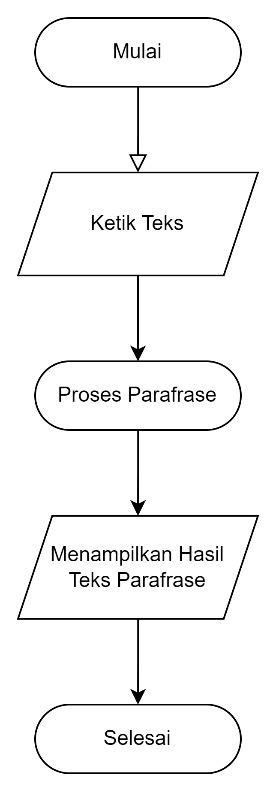
Penelitian ini menggunakan model *GPT-2* versi besar yang telah dilatih sebelumnya, kemudian dilakukan *fine-tuning* dengan dataset parafrase Bahasa Indonesia. Dataset diperoleh dari sumber publik dan diproses melalui tahapan *tokenization*, *cleaning*, dan *padding*. Model diimplementasikan menggunakan pustaka *Transformers* dari *Hugging Face* dan diintegrasikan ke aplikasi berbasis web menggunakan *Flask*.

Adapun Langkah-langkah dalam proses perancangan sistem adalah sebagai berikut :

1. 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*.
2. 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.
   3. *Model Processing*: Menggunakan pustaka *Transformer* dari *Hugging Face* untuk memproses teks input dan menghasilkan parafrase.
3. Desain Alur Kerja

a. *Flowchart*

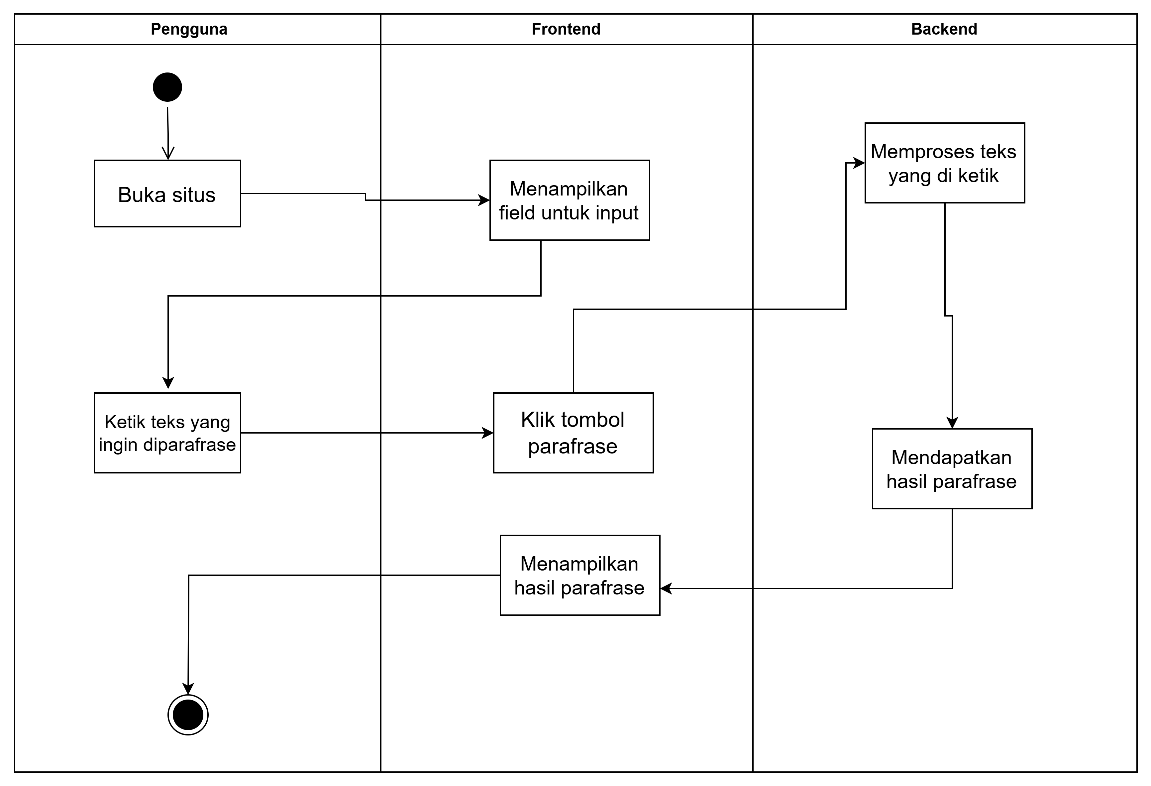
*Flowchart* ini menggambarkan proses dari input pengguna hingga keluarnya hasil parafrase:

******

**Gambar Flowchart Parafrase**

1. *Activity Diagram*

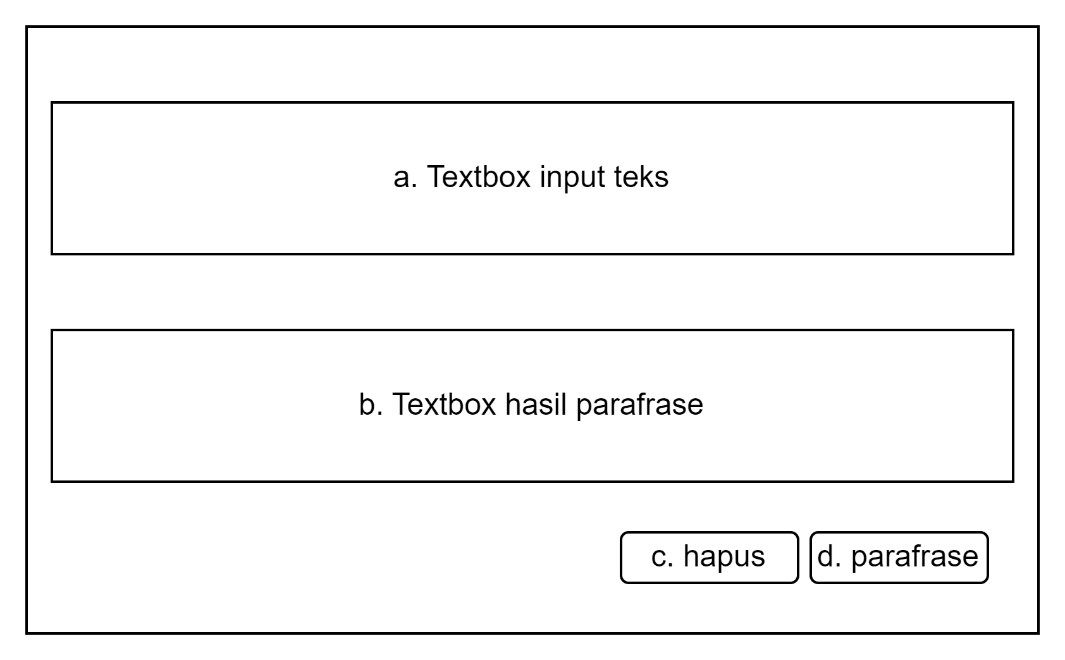
***Activity diagram* ini menggambarkan interaksi antara pengguna, frontend, dan backend:**

******

**Gambar *Activity Diagram***

* 1. Pengguna memasukkan teks yang akan diparafrase melalui *form* di halaman web.
  2. Teks dikirim ke *backend* melalui *API*.
  3. *Backend* melakukan tokenisasi dan pemrosesan teks menggunakan model *IndoT5-base-parafrase* dan menghasilkan hasil parafrase.
  4. Hasil parafrase dikirim kembali ke *frontend* dan ditampilkan kepada pengguna.

1. 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.

**Tampilan Utama**

Keterangan:

a. *Textbox* yang disediakan untuk mengisi teks yang ingin di parafrase.

b. Hasil parafrase akan muncul di *textbox* ini.

c. Menghapus seluruh teks yang telah diketik.

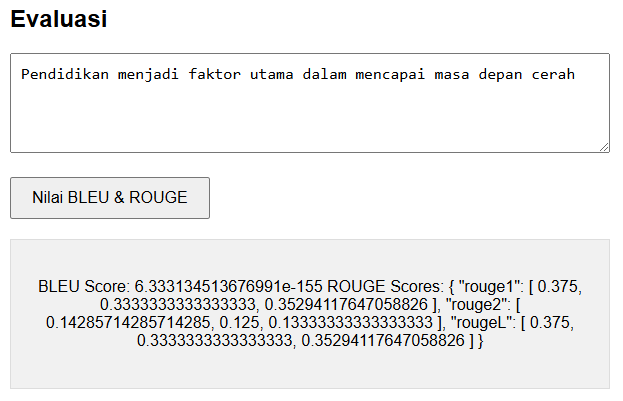
d. Tombol yang berfungsi untuk konfirmasi parafrase dari *textbox* a.

# Hasil dan Pembahasan

Setelah dilakukan implementasi maka hasilnya seperti gambar dibawah ini



**Gambar Tampilan Utama**

****

**Gambar Tampilan Hasil Evaluasi**

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

Hasil evaluasi terhadap model Transformer dalam menghasilkan parafrase teks ditunjukkan pada Tabel 4.1 berikut:

**Tabel 4.1** Hasil Nilai Rata – Rata Evaluasi

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | BLEU | ROUGE-1 | ROUGE-2 | ROUGE-L |
| GPT-2 | 50.1 | 63.4 | 48.2 | 61.7 |

Selain itu, contoh hasil parafrase dari masing-masing model ditampilkan pada Tabel 4.2 berikut:

**Tabel 4.2** 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 menjaga kesehatan tubuh dan pikiran." | "Rutin berolahraga berperan penting dalam menjaga kebugaran fisik dan mental." | "Kesehatan tubuh dan pikiran dapat dijaga dengan rutin berolahraga." | 9.6 | 59.8 |

Berdasarkan hasil tersebut, model GPT-2 yang telah difine-tuned menunjukkan performa terbaik dibandingkan model lainnya 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:

1. Model IndoT5 yang Dituning dengan *GPT-2*
2. Model menunjukkan hasil terbaik dengan nilai *BLEU* sebesar 50.1 dan *ROUGE-L* sebesar 61.7.
3. Penggunaan metrik *BLEU* dan *ROUGE* terbukti efektif dalam mengevaluasi kualitas teks parafrase.

Untuk Saran:

1. Evaluasi Tambahan Menggunakan Metrik Lain seperti METEOR atau BERTScore
2. Penggunaan Dataset yang Lebih Luas dan Bervariasi
3. Optimalisasi Model melalui *Fine-Tuning* Lebih Lanjut
4. Pengembangan Aplikasi Berbasis Web yang Lebih Interaktif

# Daftar Pustaka

[1] I. N. Purnama and N. N. Widya Utami, “Implementasi Peringkas Dokumen Berbahasa Indonesia Menggunakan Metode Text To Text Transfer Transformer (T5),” *J. Teknol. Inf. dan Komput.*, vol. 9, p. 4, 2023.

[2] A. Wahyudi, A. Pamuji, and R. Martuti, “Pelatihan Peningkatan Kemampuan Parafrasa Mahasiswa Program Studi Ilmu Keperawatan STIK Bina Husada Palembang Dengan Menggunakan Sinonim Dan Mengubah Struktur Teks,” *CARE J. Pengabdi. Multi Disiplin*, vol. 1, no. 2, pp. 30–36, 2023, doi: 10.31004/care.v1i2.15402.

[3] K. Nisa, E. Syafitri, S. R. D. Saragih, Y. Aryni, and E. Rahmadani, “Penggunaan Aplikasi Parafrasa untuk Menurunkan Plagiarisme pada Mahasiswa FKIP Universitas Asahan dalam Menyelesaikan Skripsi,” *J. War. Pengabdi. Andalas*, vol. 29, no. 1, pp. 55–60, 2022, doi: 10.25077/jwa.29.1.55-60.2022.

[4] R. Delima, A. Rachmat, and C. #2, “Otomatisasi Pembentukan Class Diagram dengan Pendekatan Metode Pemrosesan Teks dan Algoritma CombineTF,” *J. Edukasi dan Penelit. Infomatika ( JEPIN )*, vol. 10, no. 1, pp. 120–127, 2024.

[5] R. Mas, R. W. Panca, K. Atmaja1, and W. Yustanti2, “Analisis Sentimen Customer Review Aplikasi Ruang Guru dengan Metode BERT (Bidirectional Encoder Representations from Transformers),” *Jeisbi*, vol. 02, no. 3, pp. 55–62, 2021.

[6] H. Mardiana, Y. Chandra, and A. Arnelli, “Pemberdayaan Penelitian: Pentingnya Parafrase Dalam Penyusunan Artikel Jurnal Ilmiah,” *J. Pengabdi. Kolaborasi dan Inov. IPTEKS*, vol. 2, no. 4, pp. 1109–1118, 2024, doi: 10.59407/jpki2.v2i4.996.

[7] C. A. RIZQY, V. R. H. P. , NADIYA R. ALI, A. A. Salsabila, and N. D. A. , Dwi A. Putri, “EFEKTIVITAS WEB PARAFRASE ONLINE DALAM PENYUSUNAN KARYA ILMIAH/LAPORAN MAHASISWA JURUSAN TEKNOLOGI PANGAN,” *Argopuro J. Multidisiplin Ilmu Bhs.*, vol. 1, no. 4, pp. 1–14, 2023, [Online]. Available: https://ejournal.warunayama.org/index.php/argopuro/article/view/1107.

[8] S. R. Wulandari, D. W. Hamka, R. Kurniawan, and ..., “Pelatihan Teknik Parafrasa Dan Penggunaan Aplikasi Mendeley Untuk Sitasi Karya Akademik,” *J. Pengabdi. …*, vol. 2, no. 2, pp. 47–51, 2024, [Online]. Available: https://mand-ycmm.org/index.php/jpmm/article/view/693%0Ahttps://mand-ycmm.org/index.php/jpmm/article/download/693/749.

[9] C. Angga Marcelio *et al.*, “Aplikasi Analisis Wajah, Klasifikasi Gender dan Prediksi Usia Menggunakan Deep Learning pada Dataset Citra Wajah Manusia P-Issn,” *J. Media Infotama*, vol. 20, no. 1, pp. 2–6, 2024.

[10] V. Karenina, M. F. Erinsyah, and D. S. Wibowo, “Klasifikasi Rentang Usia Dan Gender Dengan Deteksi Suara Menggunakan Metode Deep Learning Algoritma Cnn (Convolutional Neural Network),” *Komputika J. Sist. Komput.*, vol. 12, no. 2, pp. 75–82, 2023, doi: 10.34010/komputika.v12i2.10516.

[11] M. Ali, A. Diwan, and D. Kumar, “Attendance System Optimization through Deep Learning Face Recognition,” *Int. J. Comput. Digit. Syst.*, vol. 15, no. 1, pp. 1527–1540, 2024, doi: 10.12785/ijcds/1501108.

[12] I. Ashish Vaswani, Ariful *et al.*, “Attention Is All You Need,” *Int. Conf. Inf. Knowl. Manag. Proc.*, no. Nips, pp. 4752–4758, 2023, doi: 10.1145/3583780.3615497.

[13] A. D. Rendragraha, M. A. Bijaksana, and A. Romadhony, “Pendekatan Metode Transformers untuk Deteksi Bahasa Kasar dalam Komentar Berita Online Indonesia,” *e-Proceeding Eng.*, vol. 8, no. 2, pp. 3385–3395, 2021.

[14] J. U. S. Lazuardi and A. Juarna, “Analisis Sentimen Ulasan Pengguna Aplikasi Joox Pada Android Menggunakan Metode Bidirectional Encoder Representation From Transformer (Bert),” *J. Ilm. Inform. Komput.*, vol. 28, no. 3, pp. 251–260, 2023, doi: 10.35760/ik.2023.v28i3.10090.

[15] A. Mukti, A. D. Hadiyanti, A. Nurlaela, and J. Panjaitan, “Sistem Analisa Sentiment Bakal Calon Presiden 2024 Menggunakan Metode NLP Berbasis Web,” *Jurikom*, vol. 6, no. 1, p. p-ISSN, 2023.

[16] M. Amien, “SEJARAH DAN PERKEMBANGAN TEKNIK NATURAL LANGUAGE PROCESSING (NLP) BAHASA INDONESIA: TINJAUAN TENTANG SEJARAH, PERKEMBANGAN TEKNOLOGI, DAN APLIKASI NLP DALAM BAHASA INDONESIA,” pp. 1–7, 2022.

[17] R. Bangun, C. Informasi, and L. Pekerjaan, “Rancang Bangun Chatbot Informasi Lowongan Pekerjaan Berbasis Whatsapp dengan Metode NLP ( Natural Language Processing ),” *BRILIANT J. Ris. dan Konseptual*, vol. 5, no. 1, pp. 619–626, 2020.

[18] I. Pendidikan, “PERAN GURU DAN SISWA DALAM MENGGUNAAN CHAT GPT (GENERATIVE PRE-TRAINING TRANSFORMER) DALAM IMPLEMENTASI PENDIDIKAN,” *KRIDA CENDEKIA*, vol. 3, no. 1, 2024.

[19] A. Aruna, E. P. Surya, A. R. Prasetyo, G. A. Wijaya, and U. N. Malang, “Integrasi Text-to-Image dan Generative Pre-Trained Transformer dalam Pengembangan Wisata Edukasi Ramah Anak Usia Dini Alby,” *VISA J. VISIONS IDEAS*, vol. 4, no. 3, pp. 2186–2198, 2024.

[20] D. Setiawan, “Peran Chat Gpt ( Generative Pre-Training Transformer ) Dalam Implementasi Ditinjau Dari Dataset,” *Innov. J. Soc. Sci. Res.*, vol. 3, no. 3, pp. 9527–9539, 2023.

[21] M. Yusnan and A. Rezki, “Pelatihan Penggunaan Alat Parafrase Online Quillbot Sebagai Upaya Peningkatan Kualitas Skripsi,” *Termasyhur J. Pengabdi. …*, vol. 2, no. 2, pp. 35–43, 2024, [Online]. Available: http://www.jurnal-umbuton.ac.id/index.php/termasyhur/article/view/5978.

[22] A. Dalam, R.- Karp, and M. Tf-idf, “Aplikasi pendeteksi plagiarisme menggunakan kombinasi metode tf-idf dan algoritma rabin-karp,” pp. 1–10, 2020.