
IMPLEMENTASI REINFORCEMENT LEARNING UNTUK SUMMARIZATION PADA ARTIKEL BERITA DI INDONESIA MENGGUNAKAN MODEL TRANSFORMER

Muhammad Alif Nasrulloh¹, Christian Sri Kusuma Aditya²
Universitas Muhammadiyah Malang, Malang

email: ¹alifnasrullohjbg@webmail.umm.ac.id, ²christianskaditya@umm.ac.id

Abstract: *This research aims to develop a Transformer-based text summarization model for Indonesian news articles and implement Reinforcement Learning to enhance summary quality. The research methodology involved using the Liputan6 dataset (Canonical subset), applying a T5-base model pre-trained with Supervised Fine-Tuning (SFT) on the same dataset, and subsequently optimizing it using the Reinforcement Learning algorithm Proximal Policy Optimization (PPO). Model performance was evaluated using the ROUGE-L metric. The results showed that the implementation of RL successfully increased the ROUGE-L F-measure score, from 0.314801 for the SFT-only model to 0.345324 for the SFT model with RL. This increase indicates an improvement in the longest common subsequence (LCS) similarity between the model-generated summaries and the reference summaries after optimization with RL. However, qualitative analysis of the generated summaries found that the RL model's summaries tended to be very short (often just one sentence) and omitted some important information present in the original articles. This suggests a limitation of the ROUGE-L metric, which focuses on lexical overlap, in fully capturing the semantic quality and completeness of the summary.*

Keyword: *Reinforcement Learning; Transformer; Abstractive Summarization; Indonesian News*

Abstrak: Penelitian ini bertujuan untuk mengembangkan model peringkasan teks berbasis *Transformer* untuk artikel berita berbahasa Indonesia dan mengimplementasikan *Reinforcement Learning* guna meningkatkan kualitas ringkasan. Metodologi penelitian melibatkan penggunaan *dataset* Liputan6 (*subset Canonical*), penerapan model T5-base yang telah dilakukan *Supervised Fine-Tuning* (SFT) pada *dataset* yang sama, kemudian dioptimalkan menggunakan algoritma *Reinforcement Learning, Proximal Policy Optimization* (PPO). Evaluasi performa model dilakukan menggunakan metrik ROUGE-L. Hasil penelitian menunjukkan bahwa implementasi RL berhasil meningkatkan skor ROUGE-L F-measure, dari 0.314801 pada model SFT saja menjadi 0.345324 pada model SFT dengan RL. Peningkatan ini mengindikasikan adanya perbaikan dalam kesamaan urutan kata terpanjang (LCS) antara ringkasan hasil model dan ringkasan referensi setelah dioptimalkan dengan RL. Namun, analisis kualitatif terhadap ringkasan yang dihasilkan menemukan bahwa ringkasan hasil model RL cenderung sangat pendek (sering kali hanya satu kalimat) dan mengabaikan beberapa informasi penting yang ada di artikel asli. Hal ini menunjukkan adanya keterbatasan metrik ROUGE-L yang berfokus pada kesamaan leksikal dalam menangkap kualitas semantik dan kelengkapan ringkasan secara keseluruhan.

Kata kunci: Reinforcement Learning; Transformer; Peringkasan Abstraktif; Berita Bahasa Indonesia

PENDAHULUAN

Era digital saat ini ditandai dengan peningkatan aksesibilitas informasi dan produksi artikel berita yang melimpah setiap harinya. Artikel ilmiah dan berita sering kali panjang dan padat informasi, membuat pembaca sulit untuk menyerap inti sarinya dengan cepat. Peringkasan teks otomatis (*automatic text summarization*) hadir sebagai solusi dengan memanfaatkan teknologi untuk meringkas teks menjadi versi singkat yang mudah dipahami (Maulidia Sari & Siti Fatonah, 2021). Pendekatan *abstractive summarization* semakin populer karena kemampuannya menghasilkan ringkasan yang menyerupai hasil olahan manusia, tidak terbatas pada penggunaan kata-kata dari teks asli seperti metode *extractive*. Model *abstractive* memungkinkan peringkasan yang lebih natural, ringkas, padat, mudah dibaca, dan kohesif dengan meniru cara manusia meringkas, yaitu memahami makna dan kemudian merangkai kalimat baru yang mengandung informasi penting (Alfhi Saputra, 2021). Perkembangan teknologi *deep learning* dan *natural language processing* (NLP) telah mendorong kemajuan signifikan dalam *abstractive summarization*, di mana Model *Transformer* menunjukkan potensi besar dalam menghasilkan ringkasan berkualitas tinggi yang mendekati ringkasan buatan manusia. Kemampuan model *Transformer* untuk memahami konteks dan hubungan antar kata melalui mekanisme *self-attention* menjadikannya pilihan tepat untuk meringkas artikel berita Indonesia (Raihanunnisa et al., 2023). Penelitian yang dilakukan oleh (Itsaini et al., 2023) dan (Khasanah & Hayaty, 2023) telah membuktikan keefektifan model *Transformer*, seperti T5 dan GPT-2, dalam meringkas artikel berita berbahasa Indonesia. Untuk lebih mengoptimalkan strategi peringkasan model *Transformer* agar menghasilkan ringkasan yang lebih akurat, ringkas, dan relevan, *Reinforcement Learning* (RL) dapat digunakan. RL memungkinkan

model untuk belajar dari *feedback* yang diberikan terhadap ringkasan yang dihasilkan, menerima *reward* untuk ringkasan berkualitas tinggi dan *penalty* untuk ringkasan yang buruk (Keneshloo et al., 2019). Pendekatan ini berfokus pada meminimalkan *negative expected reward* dan memungkinkan model berlatih berdasarkan metrik yang digunakan selama *decoding*, seperti ROUGE.

Meskipun penelitian *text summarization* sudah banyak dilakukan, fokus pada penerapan RL dengan model *Transformer* untuk meringkas artikel berita Indonesia masih terbatas. Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengembangkan model *summarization* berbasis *Transformer* untuk artikel berita dalam bahasa Indonesia, mengimplementasikan *Reinforcement Learning* guna meningkatkan kualitas peringkasan teks yang lebih akurat dan sesuai konteks, serta mengevaluasi performa model yang diimplementasikan dengan menggunakan metrik evaluasi standar seperti ROUGE.

METODE

Data Collection

Data yang digunakan dalam penelitian ini adalah *dataset* Liputan6, sebuah *dataset summarization* Bahasa Indonesia berskala besar yang diperkenalkan oleh Fajri Koto dkk. (Koto et al., 2020). *Dataset* ini dikumpulkan melalui metode web *scraping* artikel berita dari portal Liputan6.com selama sepuluh tahun (Oktober 2000 - Oktober 2010). Liputan6.com dipilih karena merupakan portal berita terkemuka di Indonesia.

Dataset Liputan6 memiliki dua varian: Canonical dan Xtreme. Penelitian ini menggunakan *subset* Canonical, karena varian ini mencakup variasi tingkat abstraksi yang lebih luas, sesuai dengan tujuan pengembangan model *abstractive summarization*, dibandingkan

set Xtreme yang berfokus pada ringkasan sangat abstrak (minimal 90% 4-gram). Untuk *sample dataset* dapat dilihat pada Tabel 1.

Pre-Processing

Tahap ini mencakup *case folding* untuk mengubah semua teks menjadi huruf kecil. Selanjutnya, dilakukan pembersihan teks (*cleaning*) menggunakan ekspresi reguler untuk menghapus karakter yang tidak relevan seperti simbol atau tanda baca yang tidak bermakna dalam konteks ringkasan. Pada tahap ini sengaja tidak melakukan langkah agresif seperti *stemming* atau menghapus seluruh *stopwords*. Hal ini dikarenakan sifat *abstractive summarization* yang memerlukan informasi lengkap dari teks asli untuk

merekonstruksi teks dalam bahasa alami, *pre-processing* yang terlalu agresif dapat mengganggu konteks dan makna, menyebabkan model kehilangan informasi penting.

Data Splitting

Dataset Liputan6 Canonical memiliki distribusi data asli sebanyak 193.883 data untuk *train*, 10.972 data untuk *development*, dan 10.972 data untuk *test*. Karena keterbatasan sumber daya komputasi seperti kapasitas penyimpanan dan waktu pemrosesan, penelitian ini hanya menggunakan sebagian kecil data dari setiap *subset*. Sebanyak 10% dari masing-masing *subset* (*train*, *development*, dan *test*) dipilih untuk mewakili keseluruhan *dataset*,

Tabel 1. Sample Data

Artikel	Ringkasan
liputan6.com, jakarta: yayasan lembaga konsumen indonesia menyetujui tindakan pemerintah menarik sejumlah produk minuman suplemen dari pasaran. produk yang ditarik itu adalah kratingdaeng, galin bugar, dan m-150 ...	kunjungan presiden ke papua jangan menjadi agenda utama. sebab, yang terpenting adalah melaksanakan semua komitmen dan kebijaksanaan pemerintah secara konkret.
liputan6.com, jakarta: meski sulit dicapai, pemerintah ingin menyelesaikan masalah aceh sebelum peringatan proklamasi 17 agustus mendatang. demikian diungkapkan menteri koordinator bidang politik, sosial, dan keamanan agum ...	pemberlakuan kenaikan tarif telepon sebesar 21, 63 persen diundur 10 juni mendatang. kenaikan harga telepon kali ini bakal diikuti sejumlah inovasi.
liputan6.com, bandung: posisi panglima tni adalah jabatan yang dapat diisi perwira tinggi berbintang tiga dan empat, dari seluruh angkatan yang ada di tni. selanjutnya, lembaga berwenang menentukan sosok yang ...	akibat pernyataan menko kesra jusuf kalla tentang dana rapel gaji guru, banyak kepala daerah didemo. padahal, dana tersebut tidak pernah ada. apkasi melayangkan surat protes.

dengan tetap memperhatikan distribusi yang seimbang. Dengan demikian, *dataset* yang digunakan terdiri dari 19.388 data untuk *train*, 1.097 data untuk *development*, dan 1.097 data untuk *test*.

Model Transformers

Model *Transformer* yang digunakan dalam penelitian ini adalah

T5-base. Model ini sudah terlebih dahulu melalui proses *Supervised Fine-Tuning* (SFT) pada *dataset* yang sama. *Base-model* ini diambil dari repositori Huggingface dengan ID “LazarusNLP/IndoNanoT5-base-Liputan6-Canonical”. T5 merupakan model yang dilatih dengan format *text-to-text* untuk berbagai tugas NLP. Varian T5-base memiliki 220 juta

parameter, yang dianggap lebih efisien untuk pelatihan dengan spesifikasi perangkat keras terbatas (Itsnaini et al., 2023). Arsitektur T5 menggunakan mekanisme *self-attention* untuk menghitung representasi *input* dan *output* secara otomatis, memungkinkannya menangkap hubungan jangka panjang dalam teks. Model SFT ini berfungsi sebagai agen RL untuk menghasilkan ringkasan.

Reinforcement Learning

Algoritma *Reinforcement Learning* yang diimplementasikan adalah PPO (*Proximal Policy Optimization*). PPO digunakan untuk mengoptimalkan kebijakan (*policy*) model SFT berdasarkan *reward* (Stienon et al., 2020). *Reward* yang digunakan adalah skor ROUGE-L, yang dihitung antara ringkasan hasil model dan *reference summary*. Skor ROUGE-L ini kemudian di-*scale* dengan dikalikan faktor 32. Faktor *scaling* ini penting untuk menghasilkan sinyal pembelajaran yang cukup kuat, mengingat nilai ROUGE-L biasanya kecil (0-1). Skor *reward* yang sudah di-*scale* ini diteruskan ke model dan diolah melalui fungsi *loss* PPO.

Fungsi *loss* PPO terdiri dari tiga komponen utama, yaitu *policy loss* yang bertujuan memaksimalkan *reward* sambil membatasi perubahan kebijakan yang drastis (dihitung menggunakan persamaan (1)), *value loss* yang bertujuan meminimalkan kesalahan prediksi nilai (dihitung menggunakan persamaan (2)) dan *entropy loss* yang bertujuan mendorong eksplorasi dalam proses pembelajaran (dihitung menggunakan persamaan (3)). Dengan demikian, fungsi *loss* PPO total dihitung dengan persamaan (4). Sinyal pembelajaran dari fungsi *loss* ini digunakan dalam proses *backpropagation* untuk memperbarui parameter model.

$$L^{\text{CLIP}} = E_t [\min(r_t(\theta) \widehat{A}_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon) \widehat{A}_t)] \quad (1)$$

Dimana $r_t(\theta)$ adalah rasio probabilitas aksi antara kebijakan baru dan lama, \widehat{A}_t adalah estimasi *advantage*, yang mengukur seberapa baik suatu aksi dibandingkan rata-rata, dan ϵ adalah *hyperparameter clipping*.

$$L^{\text{VF}} = E_t [\max((V_t - R_t)^2, (\text{clip}(V_t - R_t)^2))] \quad (2)$$

Dimana E_t adalah rata-rata kesalahan di seluruh langkah waktu, V_t adalah prediksi nilai dari model, R_t adalah *return* aktual (target), dan $(\text{clip}(V_t) - R_t)^2$ adalah *error* dengan *clipping*.

$$L^{\text{ENT}} = E_t [S[\pi_\theta](s_t)] \quad (3)$$

Dimana $S[\pi_\theta](s_t)$ adalah *entropy* dari distribusi probabilitas aksi.

$$L^{\text{PPO}} = L^{\text{CLIP}} - c_1 L^{\text{VF}} + c_2 L^{\text{ENT}} \quad (4)$$

Dimana c_1 adalah bobot untuk *value loss* dan c_2 adalah bobot untuk *entropy loss*.

Evaluasi Model

Tahap evaluasi model menggunakan metrik ROUGE-L. Metrik ini dipilih karena kemampuannya mengukur kesamaan urutan terpanjang (LCS / *Longest Common Subsequence*) antara ringkasan yang dihasilkan model dan ringkasan referensi. ROUGE-L dianggap lebih sensitif terhadap struktur kalimat dan urutan kata, serta dapat menangkap kesamaan semantik lebih baik dibandingkan ROUGE-1 dan ROUGE-2. Perhitungan skor ROUGE-L F-measure dapat dilihat pada persamaan (5) berikut.

$$\text{ROUGE} - L = \frac{2 \times \text{LCS}(S_{\text{ref}}, S_{\text{gen}})}{\text{Length}(S_{\text{ref}}) + \text{Length}(S_{\text{gen}})} \quad (5)$$

Dimana $\text{LCS}(S_{\text{ref}}, S_{\text{gen}})$ adalah panjang urutan LCS antara *true summary* / data ringkasan yang asli dan hasil ringkasan, $\text{Length}(S_{\text{ref}})$ adalah panjang *true summary* / data ringkasan yang asli, dan $\text{Length}(S_{\text{gen}})$ adalah panjang hasil ringkasan.

HASIL DAN PEMBAHASAN

Berikut kami menyajikan temuan dari dua skenario eksperimen yang dilakukan, yaitu penggunaan model *Supervised Fine-Tuning* (SFT) saja dan model SFT yang dioptimalkan dengan *Reinforcement Learning* (RL) menggunakan algoritma PPO. Kedua skenario akan menggunakan *generation config* (konfigurasi model untuk membuat ringkasan) yang sama, sehingga kedua hasilnya dapat dibandingkan.

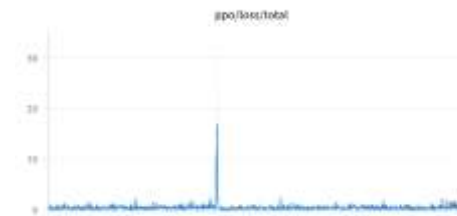
Untuk skenario pertama tidak dilakukan training pada base-model karena sudah terlatih dengan dataset yang sama. Oleh karena itu pemrosesan dataset disamakan dengan dokumentasi penyedia model sehingga dapat membandingkan hasil sebelum atau sesudah diterapkan RL.

Skenario kedua menggunakan *base-model* yang sama dengan skenario pertama. Skenario ini dilatih dengan algoritma PPO selama 1 *epoch* dengan ukuran *batch* 4 atau setara dengan 4.847 *steps* mengingat keterbatasan daya komputasi yang tersedia. Pelatihan dilakukan pada GPU P100 dari *Kaggle*

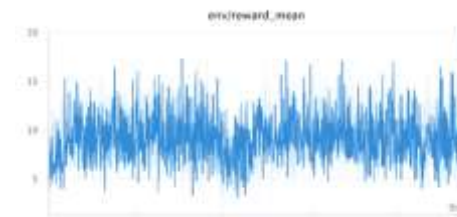
Notebook, dengan waktu pelatihan sekitar ~9 jam untuk 1 *epoch*. *Random seed* diatur pada nilai 42 untuk memastikan *reproducibility* pelatihan. Selama proses pelatihan, semua log / metrik akan di catat per step mulai dari *loss*, *reward* dan *advantage* yang diterima.



Gambar 1 Log Advantages



Gambar 2 Log Loss Total



Gambar 3 Log Reward



Gambar 4 Log Skor ROUGE-L

Tabel 2. Sample Hasil Kedua Skenario

<p>Artikel : liputan6.com, mataram: rencana penyerangan amerika serikat terhadap afghanistan berdampak negatif bagi para warga negara negeri paman sam. buktinya, langkah pengamanan terhadap warga as yang bekerja di tambang emas pt newmont nusa tenggara di nusatenggara barat semakin diperketat. beberapa waktu sebelumnya, teror bom molotov terjadi hanya sekitar dua kilometer dari lokasi pertambangan.</p>

kendati demikian, kegiatan penambangan di tambang emas tersebut tetap berlangsung seperti biasa. demikian pernyataan hubungan masyarakat pt newmont nusa tenggara anton susanto, baru-baru ini, di mataram. menurut anton, teror molotov itu terjadi beberapa saat setelah isu sweeping warga as mulai santer. sebuah kamar nomor 8, di bungalow surya, beneter, sekitar dua kilometer dari daerah tambang beneter port, diledakkan sebuah bom molotov. beberapa warga as memang menginap di sana. lantaran itulah, belakangan langkah penjagaan yang dilakukan pasukan perintis polda ntb diperketat, dengan cara bergantian setiap enam jam. sementara itu, pengamanan di kantor pt newmont di mataram dilakukan pasukan perintis polda ntb dan sejumlah anggota security 911. di kantor pusat pt newmont di lokasi tambang di batu hijau, benete, sumbawa, pengamanan langsung dilakukan pasukan brimob yang dilengkapi tim penjinak bom. seluruh pekerja asing yang berkantor pusat di nevada, as, termasuk pekerja asing asal as, berkerja di perusahaan yang memperkerjakan sedikitnya 140 pekerja asing. sebanyak 60 orang di antaranya adalah warga as yang tetap bekerja dan tidak mengambil cuti. (bmi/adhar hakim).

Reference Summary :

langkah pengamanan terhadap warga as yang bekerja pada tambang emas pt newmont nusa tenggara di ntb semakin diperketat. lokasi pertambangan newmont diteror bom molotov.

Generated Summary (Model SFT) :

rencana penyerangan amerika serikat terhadap afghanistan berdampak negatif bagi warga negara as yang bekerja di tambang emas pt newmont nusa tenggara di ntb. langkah pengamanan terhadap warga as semakin diperketat dengan cara bergantian setiap enam jam, termasuk pasukan brimob

Skor ROUGE-L : 0.4761904762

Generated Summary (Model SFT + RL) :

langkah pengamanan terhadap warga as yang bekerja di tambang emas pt newmont nusa tenggara semakin diperketat.

Skor ROUGE-L : 0.75

Berdasarkan tabel perbandingan diatas, didapatkan peningkatan skor ROUGE-L F-measure pada model yang diimplementasikan RL menunjukkan bahwa algoritma PPO berhasil memandu model untuk menghasilkan ringkasan yang memiliki kesamaan urutan kata terpanjang (LCS) yang lebih tinggi dengan ringkasan referensi. Mengingat RL memungkinkan model untuk belajar dari *feedback* berupa *reward* berdasarkan kualitas ringkasan yang dihasilkan, berbeda dengan pelatihan *supervised learning* yang meminimalkan *cross-entropy loss* (Keneshloo et al., 2019).

Meskipun skor ROUGE-L meningkat, analisis kualitatif mengungkapkan adanya kelemahan signifikan pada ringkasan hasil model RL. Beberapa diantaranya adalah ringkasannya menjadi terlalu pendek

dan kehilangan detail penting. Hal ini dapat dikaitkan dengan karakteristik metrik ROUGE-L itu sendiri. ROUGE-L mengukur kesamaan leksikal berdasarkan LCS dan kurang sensitif terhadap aspek kualitas ringkasan lainnya, seperti koherensi, kelancaran, dan keterbacaan. ROUGE memiliki keterbatasan, terutama dalam mengevaluasi model *abstractive summarization* yang dapat menghasilkan kata atau frasa baru yang tidak ada dalam teks asli atau referensi.

Adapun skor ROUGE-L keseluruhan data test pada tabel berikut.

Tabel 3 Skor ROUGE-L Data Test

	ROUGE-L	
	Model SFT	Model SFT + RL
Precision	0.273667	0.397297
Recall	0.378734	0.317939

F-measure	0.314801	0.345324
-----------	----------	----------

Metrik ini berfokus pada *overlap* kata atau n-gram dan mungkin tidak sepenuhnya menangkap kualitas semantik. Dalam kasus ini, model RL mungkin menemukan bahwa menghasilkan ringkasan yang sangat pendek namun memiliki *overlap* kata kunci yang tinggi dengan referensi adalah cara efektif untuk memaksimalkan *reward* ROUGE-L, meskipun informasi yang disampaikan menjadi tidak lengkap.

SIMPULAN

Penelitian ini menunjukkan bahwa implementasi *Reinforcement Learning* menggunakan algoritma PPO berhasil meningkatkan skor ROUGE-L pada model peringkasan teks berbasis *Transformer*. Model yang dilatih dengan SFT dan RL menunjukkan skor ROUGE-L F-measure 0.345324, lebih

tinggi dibandingkan model SFT saja yang mencapai 0.314801. Peningkatan ini mengindikasikan adanya perbaikan dalam kesamaan urutan kata terpanjang (LCS) antara ringkasan hasil model dan ringkasan referensi setelah dioptimalkan dengan RL.

Meskipun terjadi peningkatan pada metrik ROUGE-L, analisis kualitatif terhadap ringkasan yang dihasilkan menunjukkan adanya beberapa kelemahan. Ringkasan hasil model RL cenderung sangat pendek (sering kali hanya satu kalimat), mengorbankan kelengkapan informasi penting yang ada di artikel asli. Hal ini berbeda dengan ringkasan model SFT yang lebih panjang namun terkadang kurang koheren. ROUGE-L yang berfokus pada kesamaan leksikal melalui LCS mungkin mendorong model untuk menghasilkan ringkasan yang lebih pendek tetapi memiliki *overlap* kata yang tinggi dengan referensi demi memaksimalkan *reward*, meskipun substansi informasinya tidak lengkap.

DAFTAR PUSTAKA

- Alfhi Saputra, M. (2021). Peringkasan Teks Otomatis Bahasa Indonesia secara Abstraktif Menggunakan Metode Long Short-Term Memory. *E-Proceeding of Engineering: Vol.8, No.2 April 2021*, 8(2), 3474–3488.
- Itsnaini, Q. A., Hayaty, M., Putra, A. D., & Jabari, N. A. (2023). Abstractive Text Summarization using Pre-Trained Language Model “Text-to-Text Transfer Transformer (T5).” *ILKOM Jurnal Ilmiah*, 15(1), 124–131. <https://doi.org/10.33096/ilkom.v15i1.1532.124-131>
- Keneshloo, Y., Ramakrishnan, N., & Reddy, C. K. (2019). Deep transfer reinforcement learning for text summarization. *SIAM International Conference on Data Mining, SDM 2019*, 675–683. <https://doi.org/10.1137/1.9781611975673.76>
- Khasanah, A. N., & Hayaty, M. (2023). Abstractive-Based Automatic Text Summarization on Indonesian News Using Gpt-2. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, 10(1), 9–18. <https://doi.org/10.33330/jurtek.v10i1.2492>
- Koto, F., Lau, J. H., & Baldwin, T. (2020). *Liputan6: A Large-scale Indonesian Dataset for Text Summarization*. 1, 598–608. <http://arxiv.org/abs/2011.00679>
- Maulidia Sari, Y., & Siti Fatonah, N. (2021). JEPIN (Jurnal Edukasi dan Penelitian Informatika) Peringkasan Teks Otomatis pada Modul Pembelajaran Berbahasa Indonesia

-
- Menggunakan Metode Cross Latent Semantic Analysis (CLSA). *Jurnal Edukasi Dan Penelitian Informatika*, 7(2), 153–159. www.kompas.com.
- Raihanunnisa, F., Arhami, M., & Hidayat, R. (2023). Pendekatan Hybrid Pada Sistem Peringkat Teks Artikel Berita Bahasa Inggris Menggunakan Natural Language Processing. *Telematika Mkom*, 15(2), 86–92. <https://journal.budiluhur.ac.id/index.php/telematika/>
- Stiennon, N., Ouyang, L., Wu, J., Ziegler, D. M., Lowe, R., Voss, C., Radford, A., Amodei, D., & Christiano, P. (2020). Learning to summarize from human feedback. *Advances in Neural Information Processing Systems, 2020-Decem(NeurIPS)*, 1–14.
- Wang, M., Xie, P., Du, Y., & Hu, X. (2023). T5-Based Model for Abstractive Summarization: A Semi-Supervised Learning Approach with Consistency Loss Functions. *Applied Sciences (Switzerland)*, 13(12). <https://doi.org/10.3390/app1312711>