Teknik: Jurnal Ilmu Teknik dan Informatika
Volume 6 Nomor 2 Oktober 2026
e-ISSN: 2798-2513; p-ISSN: 2808-8751, Hal *41-52*
DOI: https://doi.org/10.51903/teknik.v6i2.1254
Available online at: https://journal.stiestekom.ac.id/index.php/TEKNIK

Analisis Perbandingan Kinerja Model Peringkasan Teks Ekstraktif dan Abstraktif (T5, mT5) pada Artikel Berita Berbahasa Indonesia

Dea Anggraini[1], Farizky Aulia Lubis[2], Callysa Elistia[3], Azis Kurniadi[4], Anggia Puteri[5]
[1-4]Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Medan

Korespondensi penulis: anggrainidea2006@gmail.com

Abstract. *The increasing number of online news articles published every day has created challenges for readers in obtaining important information efficiently. Automatic text summarization offers a solution by generating concise versions of news articles while preserving essential information. This study aims to compare the performance of extractive and abstractive text summarization models for Indonesian-language news articles, namely IndoBERT, T5-Base, and mT5 XL-Sum. The research employed a quantitative experimental approach using the IndoSum dataset, with 200 news articles from the test split selected as evaluation samples. IndoBERT was implemented as an extractive model through binary sequence classification, while T5-Base and mT5 XL-Sum were evaluated using a sequence-to-sequence generative approach. Model performance was measured using ROUGE-1, ROUGE-2, and ROUGE-L metrics to assess information coverage, syntactic quality, and sequence coherence. The results show that T5-Base achieved the highest performance with ROUGE-1 of 0.5686, ROUGE-2 of 0.4950, and ROUGE-L of 0.5365. IndoBERT obtained competitive results with ROUGE-1 of 0.4377, ROUGE-2 of 0.3291, and ROUGE-L of 0.3784, while mT5 XL-Sum produced the lowest scores among the evaluated models. These findings indicate that a monolingual abstractive model specifically trained for Indonesian language tasks is more effective than extractive and multilingual approaches for Indonesian news summarization. The study contributes empirical evidence regarding the suitability of Transformer-based summarization models and can serve as a reference for future development of automatic summarization systems for Indonesian-language content.*
Keywords: *Abstractive Summarization; Indobert; Indonesian News; MT5; Text Summarization.*

Abstrak. Jumlah artikel berita daring yang terus meningkat setiap hari menyebabkan pembaca semakin sulit memperoleh informasi penting secara cepat dan efisien. Peringkasan teks otomatis menjadi salah satu solusi yang dapat digunakan untuk menghasilkan versi ringkas dari artikel berita tanpa menghilangkan informasi utama yang terkandung di dalamnya. Penelitian ini bertujuan untuk membandingkan kinerja model peringkasan teks ekstraktif dan abstraktif pada artikel berita berbahasa Indonesia, yaitu IndoBERT, T5-Base, dan mT5 XL-Sum. Metode penelitian yang digunakan adalah eksperimen kuantitatif dengan memanfaatkan dataset IndoSum. Sebanyak 200 artikel berita dari data uji digunakan sebagai sampel evaluasi untuk seluruh model. IndoBERT diterapkan sebagai model ekstraktif melalui pendekatan klasifikasi urutan biner, sedangkan T5-Base dan mT5 XL-Sum diuji menggunakan pendekatan generatif sequence-to-sequence. Evaluasi dilakukan menggunakan metrik ROUGE-1, ROUGE-2, dan ROUGE-L untuk mengukur cakupan informasi, kualitas struktur sintaksis, serta kesesuaian urutan informasi dalam ringkasan yang dihasilkan. Hasil penelitian menunjukkan bahwa model T5-Base memperoleh performa terbaik dengan nilai ROUGE-1 sebesar 0,5686, ROUGE-2 sebesar 0,4950, dan ROUGE-L sebesar 0,5365. IndoBERT menghasilkan performa yang cukup kompetitif dengan nilai ROUGE-1 sebesar 0,4377, ROUGE-2 sebesar 0,3291, dan ROUGE-L sebesar 0,3784, sedangkan mT5 XL-Sum memperoleh nilai terendah dibandingkan model lainnya. Temuan ini menunjukkan bahwa model abstraktif monolingual yang dilatih khusus untuk bahasa Indonesia lebih efektif dalam menghasilkan ringkasan berita dibandingkan pendekatan ekstraktif maupun model multilingual. Penelitian ini diharapkan dapat menjadi referensi dalam pengembangan sistem peringkasan teks otomatis berbasis Transformer untuk dokumen berbahasa Indonesia.
Kata kunci: Abstractive Summarization; Berita Indonesia; Indobert; MT5; Peringkasan Teks.

1. LATAR BELAKANG

Diterima: 01 Juni 2026; Direvisi: 17 Juli 2026; Disetujui: 03 September 2026; Tersedia Online: 20 September 2026; Diterbitkan: 06 Oktober 2026.

Jumlah artikel berita daring yang beredar setiap harinya terus bertambah, membuat pembaca semakin kesulitan menyerap informasi secara menyeluruh dalam waktu terbatas (Gurning et al., 2025). Banyak orang akhirnya hanya membaca judul atau paragraf awal saja, padahal isi berita seringkali memuat detail yang tidak kalah penting untuk dipahami. Dari kondisi ini muncul kebutuhan nyata akan sistem yang bisa meringkas artikel secara otomatis, menghasilkan versi lebih pendek dari teks asli tanpa menghilangkan informasi utamanya (Krishna et al., 2025).

Kendati demikian, model peringkasan teks yang awalnya dioptimalkan untuk dokumen berbahasa Inggris tidak dapat serta-merta diterapkan secara instan pada bahasa lain (Itsnaini et al., 2023). Hal ini dikarenakan Bahasa Indonesia memiliki karakteristik linguistik yang sangat khas dan kompleks, di mana formasi struktur kosakatanya melibatkan operasi morfologi yang kaya seperti afiksasi, pemajemukan, serta reduplikasi kata (Wilie et al., 2020). Kompleksitas struktural tersebut menyebabkan tahapan pra-pemrosesan teks konvensional, seperti penghapusan kata fungsional (*stopword removal*) dan pemotongan kata dasar (*stemming*), secara empiris justru dapat menurunkan kinerja akurasi model bahasa berbasis Transformer karena berisiko menghilangkan token-token kontekstual esensial yang sangat dibutuhkan untuk memahami keterkaitan makna secara utuh (Khairani et al., 2024). Hambatan struktural ini menegaskan perlunya pengujian komparatif menyeluruh yang disesuaikan secara khusus dengan karakteristik sintaksis teks berita berbahasa Indonesia.

Penelitian peringkasan teks berbahasa Indonesia sebelumnya banyak bertumpu pada metode ekstraktif berbasis statistik seperti TF-IDF dan TextRank yang bekerja dengan memilih kalimat berdasarkan frekuensi kata (Gurning et al., 2025). Metode ini tidak memerlukan data pelatihan, tapi hasilnya sering terasa kaku karena terlalu bergantung pada kemunculan kata harfiah. Seiring berkembangnya deep learning, pendekatan berbasis BERT mulai banyak diuji untuk peringkasan ekstraktif teks Indonesia karena kemampuannya memahami konteks kalimat secara dua arah (Trisna et al., 2025; Wicaksono et al., 2023). Untuk pendekatan abstraktif, T5 monolingual sudah diuji pada teks berbahasa Indonesia dan menunjukkan hasil yang cukup baik, meski masih rentan terhadap redundansi dan pemotongan teks akibat batas token (Itsnaini et al., 2023; Satya et al., 2025). Model mT5 yang dilatih pada 101 bahasa termasuk Indonesia menawarkan kapasitas lintas bahasa yang lebih luas, namun di sisi lain juga berisiko

menghasilkan teks yang kurang akurat untuk konteks lokal karena data pelatihannya sangat beragam (Satya et al., 2025; Xue et al., 2021).

Sementara itu, urusan implementasi varian model T5 multilingual (mT5) yang menawarkan kapasitas parameter raksasa lintas bahasa lintas korpus global didapati masih rentan menghasilkan halusinasi teks dan kekeliruan interpretasi detail fakta lokal bahasa Indonesia karena dilatih pada data yang sangat umum (Krishna et al., 2025; Satya et al., 2025).

Meskipun studi terhadap masing-masing arsitektur tersebut telah banyak dilakukan secara terpisah, terdapat celah penelitian (*research gap*) yang nyata di mana belum ada kajian ilmiah yang secara mendalam mempertandingkan performa pendekatan ekstraktif berbasis statistik (TF-IDF + TextRank), model abstraktif monolingual (T5), dan model abstraktif multilingual (mT5) secara langsung dan bersamaan dalam satu kerangka kerja yang seragam untuk konteks bahasa Indonesia. Penelitian terdahulu umumnya berfokus secara parsial, baik hanya berpusat pada variasi representasi semantik algoritma graf ekstraktif (Gurning et al., 2025) maupun perbandingan internal di dalam kelompok model generatif sejenis (Satya et al., 2025). Oleh karena itu, kebaruan ilmiah dari penelitian ini terletak pada pengujian komparatif ketiga metodologi yang secara filosofis bertolak belakang tersebut untuk meringkas artikel berita berbahasa Indonesia dengan standarisasi metrik evaluasi ROUGE (ROUGE-1, ROUGE-2, dan ROUGE-L) guna mengukur akurasi cakupan informasi fakta, kelancaran sintaksis lokal, serta kesesuaian alur urutan tingkat kalimat secara objektif dan seimbang.

Berangkat dari perbedaan kapabilitas dan konfigurasi arsitektur dasar dari setiap model, rumusan masalah ilmiah dalam penelitian ini bermuara pada pengujian apakah perluasan kapasitas parameter berskala global mampu mengompensasi pemahaman konteks kebahasaan lokal. Berdasarkan premis pemodelan teks tersebut, penelitian ini mengajukan hipotesis tersirat bahwa pendekatan abstraktif berbasis model mT5 yang didukung transfer pengetahuan lintas bahasa (*cross-lingual transfer*) serta dilatih pada korpus multilingual masif (Xue et al., 2021), diperkirakan mampu menghasilkan teks ringkasan berita berbahasa Indonesia yang secara kualitas lebih koheren, mengalir, dan memiliki keterbacaan alami yang superior jika dibandingkan dengan performa T5 monolingual yang fokus lokal maupun kekakuan pemilihan kalimat harfiah dari pendekatan ekstraktif TF-IDF + TextRank.

Sesuai dengan rumusan hipotesis ilmiah yang diajukan, tujuan utama dari penelitian ini adalah untuk mengevaluasi dan membandingkan secara empiris performa sistem peringkasan otomatis antara pendekatan ekstraktif (TF-IDF + TextRank), model T5, dan model mT5 pada korpus dokumen berita berbahasa Indonesia. Melalui perbandingan nilai ROUGE-1, ROUGE-2, dan ROUGE-L dari ketiga model, penelitian ini bertujuan memberi gambaran nyata tentang model mana yang paling sesuai untuk peringkasan berita berbahasa Indonesia, baik dari sisi kualitas ringkasan maupun efisiensi komputasi.

2. KAJIAN TEORITIS

Peringkasan Teks Otomatis

Peringkasan teks otomatis menghasilkan versi lebih pendek dari teks asli dengan tetap mempertahankan informasi utamanya (Jangtap et al., 2024). Pendekatan extractive memilih kalimat langsung dari teks asli tanpa mengubah redaksinya, sedangkan abstractive menghasilkan kalimat baru seperti cara manusia meringkas (Krishna et al., 2025). Pada teks berbahasa Indonesia, pendekatan extractive berbasis BERT terbukti lebih baik dibanding model multibahasa seperti mBERT dan DistilBERT karena IndoBERT dilatih khusus pada korpus Indonesia (Liem et al., 2025). Trisna et al., (2025) menggabungkan sentence embedding dari IndoBERT dengan algoritma LexRank untuk meranking kalimat pada berita Indonesia dan mencatat skor ROUGE-1 tertinggi sebesar 0,70. Wicaksono et al., (2023) juga membuktikan BERT mampu meringkas dokumen teks panjang berbahasa Indonesia secara akurat menggunakan pendekatan sequence classification.

Model Transformer: T5 dan mT5

Transformer mengandalkan mekanisme self-attention untuk memahami hubungan antar kata dan menjadi dasar banyak model NLP modern termasuk untuk peringkasan teks (Jangtap et al., 2024). T5 mengubah semua tugas NLP menjadi format teks-ke-teks sehingga proses fine-tuning lebih seragam dan mudah diterapkan (Itsnaini et al., 2023). Versi multibahasanya yaitu mT5 dilatih pada 101 bahasa termasuk bahasa Indonesia sehingga lebih sesuai untuk teks non-Inggris (Xue et al., 2021). Alipour & Aydın, (2025) membuktikan mT5 menghasilkan ringkasan lebih baik dibanding model monolingual untuk bahasa dengan data pelatihan terbatas, yang kondisinya mirip dengan bahasa Indonesia.

NLP Bahasa Indonesia dan Dataset

Bahasa Indonesia punya karakteristik tersendiri sehingga model berbahasa Inggris tidak bisa langsung dipakai begitu saja (Wilie et al., 2020). Tahapan preprocessing seperti tokenisasi dan penghapusan stopword memberi pengaruh besar terhadap performa model NLP bahasa Indonesia (Khairani et al., 2024). Dewi & Widiastuti, (2022) juga mencatat bahwa kualitas dataset berbahasa Indonesia berpengaruh langsung pada hasil peringkasan, terutama untuk pendekatan abstractive. Dataset yang digunakan dalam penelitian ini adalah IndoSum, yang berisi 19.000 pasang artikel berita dan ringkasan berbahasa Indonesia.

Metrik Evaluasi dan Penelitian Terdahulu

ROUGE mengukur kualitas ringkasan dengan menghitung tumpang tindih antara output model dan ringkasan referensi, mencakup ROUGE-1, ROUGE-2, dan ROUGE-L (Barbella & Tortora, 2022). Metrik ini lebih andal untuk pendekatan extractive karena abstractive cenderung memakai kata berbeda dari referensi meski maknanya sama (Citarella et al., 2025). Khasanah & Hayaty, (2023) menggunakan GPT-2 untuk peringkasan berita Indonesia dan mendapat skor ROUGE yang kompetitif, sementara Hartawan et al., (2024) dengan BART pada Liputan6 mencatat hasil lebih baik dari model berbasis RNN. Dari penelitian-penelitian tersebut terlihat belum ada yang membandingkan langsung pendekatan extractive berbasis IndoBERT, T5, dan mT5 dalam satu eksperimen terkontrol pada data berita Indonesia, dan penelitian ini mengisi celah tersebut.

3. METODE PENELITIAN

Penelitian eksperimental kuantitatif ini membandingkan efektivitas tiga arsitektur Transformer dalam meringkas teks Berbahasa Indonesia. Populasi data bersumber dari korpus publik IndoSum yang memuat artikel berita beserta ringkasan referensinya. Sampel penelitian diambil secara sekuensial dari komponen *test-split* sebanyak 200 artikel berita pertama. Ukuran sampel ini diterapkan secara konsisten pada seluruh model untuk menjaga keadilan komparasi dan memitigasi kendala beban komputasi pemrosesan generatif.

Model ekstraktif pada penelitian ini menggunakan arsitektur IndoBERT (*indobenchmark/indobert-base-p1*) yang melalui tahap pelatihan lanjutan (*fine-tuning*) sebagai pengklasifikasi urutan biner (*binary sequence classifier*) pada tingkat kalimat. Proses tersebut menggunakan dokumen dari himpunan data IndoSum, dengan

memanfaatkan berkas train.01.jsonl sebagai data pelatihan (*training set*) dan berkas dev.01.jsonl sebagai data validasi (*validation set*). Sebelum pelatihan dimulai, dilakukan pembuatan label target (*oracle labels*) menggunakan teknik tumpang-tindih kata (*word overlap*) antara kalimat pada artikel asli dengan ringkasan acuan manusia. Sebuah kalimat diberi label 1 jika memiliki nilai *recall* lebih dari 30% atau minimal memiliki kesamaan 3 kata dengan ringkasan acuan, sedangkan label 0 diberikan jika kalimat tidak memenuhi kriteria tersebut. Pelatihan model dilakukan selama 3 *epoch* dengan konstanta *learning rate* sebesar 2e-5 dan ukuran *batch* (*batch size*) sebesar 16.

Berbeda dengan model ekstraktif, model abstraktif (T5 dan mT5) langsung diuji secara generatif (*Sequence-to-Sequence*) menggunakan *checkpoint* model praterlatih. Data sekunder dari berkas JSONL diekstraksi menggunakan skrip Python yang dilengkapi fungsi validasi tipe data (*type-checking*) untuk mencegah distorsi karakter teks tunggal selama pemuatan kalimat. Karakteristik teknis, metode pelatihan, dan parameter inferensi ketiga model dirangkum dalam Tabel 1.

MODEL	**IndoBERT (Ekstraktif)**	**T5-Base (Abstraktif)**	**mT5 XL-Sum (Abstraktif)**
Arsitektur Datar	indobert-base-p1	cahya/t5-base-indonesian	csebuetnlp/mT5_multilingual
Metode Pelatihan	*Fine-tuning biner (3 epoch, LR 2e-5, batch 16)*	Evaluasi langsung via checkpoint	Evaluasi langsung via checkpoint
Teknik Processing	*Pemisahan kalimat (sentence tokens)*	*Chunking* (400 kata, 50 *overlap*)	*Chunking* (400 kata, 50 *overlap*)
Algoritma Inferensi	Urutan Probabilitas Klasifikasi Tertinggi	*Beam Search (beams=8, ngram=3)*	*Beam Search (beams=8, ngram=3)*
Panjang Output	Rasio kompresi 30% kalimat asli	Dinamis (150 token baru/chunk)	Dinamis (150 token baru/chunk)

Alat analisis data yang digunakan untuk mengevaluasi kualitas luaran ringkasan dari ketiga model penelitian adalah metrik ROUGE (Recall-Oriented Understudy for Gisting Evaluation). Performa akhir dari setiap model dievaluasi berdasarkan nilai rata-rata (mean) F1-Score yang diperoleh dari 200 artikel uji. Ketiga indikator utama yang menjadi tolak ukur pengujian ini dirangkum pada Tabel 2.

Metrik evaluasi	**Objek Pengukuran Teks**	**Representasi Kualiatas Ringkasan**
ROUGE-1	Tumpang-tindih unit kata tunggal (*unigram*)	Mengukur akurasi cakupan informasi dan keakuratan fakta.

ROUGE-2	Tumpang-tindih pasangan kata berurutan (*bigram*)	Menilai kelancaran struktur sintaksis dan tata bahasa lokal.
ROUGE-L	Substring terpanjang yang sama (*Longest Common Subsequence*)	Menilai kesesuaian alur dan urutan informasi pada tingkat kalimat.

4. HASIL DAN PEMBAHASAN

Proses Pengumpulan dan Eksekusi Data

Lingkungan Eksperimen dan Durasi Komputasi

Proses pengumpulan data luaran berupa ringkasan teks dan pengujian performa model dilakukan sepenuhnya menggunakan platform cloud computing Google Colaboratory. Pengujian ini menggunakan sampel acak sebanyak 200 artikel dari data uji (test-split) pada dataset IndoSum dengan tujuan menjaga konsistensi evaluasi. Untuk memberikan gambaran yang jelas mengenai perbedaan infrastruktur dan efisiensi waktu yang dibutuhkan oleh masing-masing arsitektur, rincian teknis eksekusi dirangkum dalam Tabel 3.

Tabel 3. Spesifikasi Perangkat Eksperimen dan Durasi Komputasi Model

Arsitektur Model	Jenis Pendekatan	Jenis Perangkat	Durasi Eksekusi (200 Artikel)
IndoBERT	Ekstraktif	CPU *Default* Colab	8 – 10 Menit
T5-Base	Abstraktif	GPU T4 Gratis	± 18 Menit (Evaluasi: 15 Menit)
mT5 XL-Sum	Abstraktif	GPU T4 Gratis	18 Menit 30 Detik (Evaluasi: 15 Menit)

Berdasarkan data pada Tabel 3, terlihat adanya perbedaan karakteristik yang kontras antara metode peringkasan ekstraktif dan abstraktif. Model IndoBERT yang berjalan dengan pendekatan ekstraktif mampu menyelesaikan pemrosesan 200 artikel dalam waktu singkat, yaitu sekitar 8 hingga 10 menit, meskipun hanya menggunakan kapasitas CPU standar tanpa bantuan akselerasi grafis. Model ekstraktif bekerja dengan prinsip klasifikasi biner, yaitu langsung mengambil kalimat dari teks asli tanpa proses generasi, sehingga beban komputasinya lebih ringan dibanding model abstraktif.

T5-Base dan mT5 XL-Sum membutuhkan waktu eksekusi sekitar 18 menit, hampir dua kali lebih lama dibanding IndoBERT, dengan durasi terpanjang berada pada proses evaluasi akhir yang memakan waktu 15 menit. Waktu komputasi yang relatif sama antara T5 dan mT5 ini disebabkan oleh penggunaan strategi pemotongan teks (*chunking*) serta parameter algoritma *Beam Search* yang sudah diseragamkan. Karakteristik

komputasi yang lebih lama pada T5 dan mT5 ini tidak lepas dari mekanisme generasi teksnya yang bersifat *auto-regressive*, di mana model harus memprediksi dan merangkai token kata secara sekuensial (berurutan). Karena beban komputasi yang intensif tersebut, pemanfaatan GPU T4 menjadi sangat krusial untuk mempercepat waktu pemrosesan sekaligus menghindari risiko kegagalan eksekusi (*timeout*).

Hasil Analisis Data dan Keterkaitan Konsep

Evaluasi Kuantitatif Metrik ROUGE

Keberhasilan dari masing-masing model dalam mereduksi teks berita tanpa kehilangan informasi utama diukur menggunakan metrik standardisasi ROUGE (ROUGE-1, ROUGE-2, dan ROUGE-L). Hasil komparasi nilai rata-rata (mean) dari F1-Score yang diperoleh melalui pengujian terhadap 200 artikel uji disajikan dalam Tabel 4.

Tabel 4. Perbandingan Nilai Rata-Rata ROUGE F1-Score

Arsitektur Model	Jenis Pendekatan	ROUGE-1	ROUGE-2	ROUGE-L
T5-Base	Abstraktif	0.5686	0.4950	0.5365
IndoBERT	Ekstraktif	0.4377	0.3291	0.3784
mT5 XL-Sum	Abstraktif	0.4064	0.3175	0.3724

Performa model T5-Base didukung oleh arsitekturnya yang sejak awal dilatih menggunakan korpus monolingual Bahasa Indonesia. Fokus pada satu bahasa tersebut memberikan pemahaman yang lebih spesifik terhadap struktur kalimat dan konteks berita lokal. Sebagai model abstraktif, T5-Base mampu menyusun kalimat baru (parafrasa) yang terstruktur, dengan tetap mempertahankan informasi utama agar selaras dengan ringkasan acuan manusia.

Di sisi lain, model IndoBERT menghasilkan performa yang kompetitif karena menggunakan pendekatan ekstraktif dan telah melalui proses *fine-tuning* menggunakan dataset IndoSum. Proses *fine-tuning* ini membuat IndoBERT lebih optimal dalam mengenali dan memilih kalimat utama dari artikel asli. Meskipun pendekatan ekstraktif memiliki karakteristik luaran yang cenderung kaku karena hanya menyalin kalimat utuh tanpa mengubah struktur kata, kombinasi antara hasil *fine-tuning* dan metode penyalinan langsung tersebut menjaga tingkat tumpang-tindih kata pada metrik ROUGE tetap stabil.

Dari pola tersebut, wajar jika skor ROUGE IndoBERT akhirnya lebih tinggi dari mT5 XL-Sum pada pengujian ini. Sebagai model abstraktif, mT5 XL-Sum bertugas merangkai kata demi kata dari awal. Namun, karena model ini bersifat multilingual (dilatih dalam banyak bahasa sekaligus) dan diuji tanpa melalui proses *fine-tuning*

lanjutan pada dataset IndoSum, hasil generasinya menjadi kurang optimal untuk pengujian lokal ini. Sifat model abstraktif yang memparafrasa kalimat, ditambah cakupan bahasa yang luas pada model multilingual, membuat pilihan kosakata mT5 sering kali berbeda dengan kata kunci pada ringkasan acuan. Perbedaan dalam pencocokan kata secara tepat inilah yang menyebabkan penurunan nilai skor ROUGE-2 dan ROUGE-L pada mT5 XL-Sum.

Perbedaan performa antara arsitektur monolingual dan multilingual ini sejalan dengan temuan yang dipublikasikan oleh Satya et al., (2025) dalam jurnal *Sistemasi*. Dalam studi evaluasi terhadap varian arsitektur T5, Satya et al., (2025) menyimpulkan bahwa model T5-Base Monolingual menunjukkan akurasi ROUGE yang lebih baik dibandingkan varian mT5. Selaras dengan literatur tersebut, pengujian pada dataset IndoSum ini mengonfirmasi bahwa model yang dilatih khusus menggunakan korpus Bahasa Indonesia (monolingual) memberikan hasil yang lebih optimal dibandingkan model dengan jangkauan bahasa global (multilingual) untuk tugas peringkasan teks berita lokal. Temuan ini juga didukung oleh Wiratmoko et al., (2025) yang membandingkan beberapa algoritma machine learning untuk summarization teks Indonesia dan menyimpulkan bahwa model berbasis transformer secara konsisten mengungguli pendekatan statistik konvensional. Wiradinata & Mawardi, (2025) juga menguji pendekatan RAG untuk abstractive summarization berita berbahasa Indonesia dan menemukan bahwa metode ini efektif dalam menghasilkan ringkasan yang ringkas, meski masih perlu peningkatan dalam menangkap keseluruhan konteks berita. Selaras dengan literatur tersebut, pengujian pada dataset IndoSum ini mengonfirmasi bahwa model yang dilatih khusus menggunakan korpus Bahasa Indonesia memberikan hasil yang lebih optimal dibandingkan model dengan jangkauan bahasa global untuk tugas peringkasan teks berita lokal.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa ketiga model yang diuji memiliki karakteristik dan performa yang berbeda dalam melakukan peringkasan artikel berita berbahasa Indonesia. Dari hasil evaluasi menggunakan metrik ROUGE, model T5-Base memperoleh nilai tertinggi pada seluruh aspek penilaian, yaitu ROUGE-1, ROUGE-2, dan ROUGE-L. Hasil tersebut menunjukkan bahwa model abstraktif yang dikembangkan khusus untuk Bahasa Indonesia mampu menghasilkan ringkasan yang

lebih baik dibandingkan model ekstraktif maupun model multilingual yang digunakan dalam penelitian ini.

Sementara itu, model mT5 XL-Sum memperoleh nilai yang lebih rendah dibandingkan dua model lainnya. Hal ini menunjukkan bahwa model multilingual belum tentu memberikan hasil terbaik apabila tidak disesuaikan terlebih dahulu dengan karakteristik data berbahasa Indonesia.

Berdasarkan temuan tersebut, model T5-Base dapat dipertimbangkan sebagai alternatif yang efektif untuk peringkasan otomatis pada artikel berita berbahasa Indonesia. Untuk penelitian berikutnya, pengujian dapat dilakukan dengan jumlah data yang lebih banyak, cakupan topik berita yang lebih beragam, serta proses fine-tuning pada model multilingual agar performanya dapat ditingkatkan dan menghasilkan ringkasan yang lebih sesuai dengan konteks bahasa Indonesia. Keterbatasan penelitian ini mencakup jumlah sampel yang hanya 200 artikel dan tidak dilakukannya fine-tuning pada model mT5, sehingga perbandingan antar model belum sepenuhnya setara.

DAFTAR REFERENSI

Alipour, N., & Aydın, S. (2025). Abstractive summarization using multilingual text-to-text transfer transformer for the Turkish text. *IAES International Journal of Artificial Intelligence (IJ-AI)*, *14*(2), 1587–1596. https://doi.org/10.11591/ijai.v14.i2.pp1587-1596

Barbella, M., & Tortora, G. (2022). Rouge Metric Evaluation for Text Summarization Techniques. *SSRN (Social Science Research Network)*. https://doi.org/10.2139/ssrn.4120317

Citarella, A. A., Barbella, M., Ciobanu, M. G., De Marco, F., Di Biasi, L., & Tortora, G. (2025). Assessing the effectiveness of ROUGE as unbiased metric in Extractive vs. Abstractive summarization techniques. *Journal of Computational Science*, *87*. https://doi.org/10.1016/j.jocs.2025.102571

Dewi, K. E., & Widiastuti, N. I. (2022). Automatic Summarization of Indonesian Texts Using a Hybrid Approach. *Jurnal Teknologi Informasi Dan Pendidikan*, *15*(1), 37–43. https://doi.org/10.24036/jtip.v15i1.451

Gurning, Y. C., Saragih, S. C., Lase, Y. Y., & Julham. (2025). Perbandingan TextRank Berbasis TF-IDF dan Word2Vec dalam Peringkasan Teks Berita Bahasa Indonesia. *Jurnal Komputer Teknologi Informasi Sistem Informasi (JUKTISI)*, *4*(2), 922–928. https://doi.org/10.62712/juktisi.v4i2.552

Hartawan, G., Maylawati, D. S., & Uriawan, W. (2024). Bidirectional and Auto-Regressive Transformer (BART) for Indonesian Abstractive Text Summarization. *Jurnal Informatika Polinema*, *10*(4), 535–542. https://doi.org/10.33795/jip.v10i4.5242

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

Jangtap, V., Parlewar, P., Dhande, S., Langhe, A., Choundary, H., & Mishra, A. (2024). View of Advancements in Text Summarization Through Machine Learning: A Comprehensive Survey and Analysis. *Journal of Electrical Systems*, 833–849.

Khairani, U., Mutiawani, V., & Ahmadian, H. (2024). Pengaruh Tahapan Preprocessing Terhadap Model Indobert Dan Indobertweet Untuk Mendeteksi Emosi Pada Komentar Akun Berita Instagram. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIIK)*, *11*(4), 887–894. https://doi.org/10.25126/jtiik.1148315

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. http://publications.lib.chalmers.se/records/fulltext/245180/245180.pdf%0Ahttps://hdl.handle.net/20.500.12380/245180%0Ahttp://dx.doi.org/10.1016/j.jsames.2011.03.003%0Ahttps://doi.org/10.1016/j.gr.2017.08.001%0Ahttp://dx.doi.org/10.1016/j.precamres.2014.12

Krishna, K. M. R., Somasundaram, K., Arulmozhivarman, P., Immanuel, S. A., & Rajkumar, E. R. (2025). Deep learning for text summarization using NLP for automated news digest. *Scientific Reports*. https://doi.org/10.1038/s41598-025-20224-1

Kurniawan, K., & Louvan, S. (2018). IndoSum: A new benchmark dataset for Indonesian text summarization. *2018 International Conference on Asian Language Processing (IALP)*, 215–220. https://doi.org/10.1109/IALP.2018.8629109

Liem, J. V., Santoso, M. I., Pohan, S., Hasani, M. F., & Maulina, A. (2025). Extractive Indonesian News Text Summarization using DistilBERT, IndoBERT, MBERT, and RoBERTa. *Procedia Computer Science*, *269*, 1087–1096. https://doi.org/10.1016/j.procs.2025.09.050

Satya, M. W. B. D., Luthfiarta, A., & Althoff, M. N. (2025). Comparative Analysis of T5 Model Performance for Indonesian Abstractive Text Summarization. *Sistemasi: Jurnal Sistem Informasi*, *14*(3), 1092–1106. https://doi.org/10.32520/stmsi.v14i3.4884

Trisna, I. N. P., Vihikan, W. O., & Azizah, A. Z. N. (2025). Combining BERT and Graph-Based Ranking for Extractive Summarization of Indonesian News Articles. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, *11*(2), 196–202. https://doi.org/10.24014/coreit.v11i2.37929

Wicaksono, G. W., Al’asqalani, S. F., Azhar, Y., Putri Hidayah, N., & Andeawan. (2023). Automatic Summarization of Court Decision Documents Over Narcotic Cases Using BERT. *INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION*, *7*(2), 416–422. www.joiv.org/index.php/joiv

Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. (2020). IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. *In Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, 843–857. https://doi.org/10.18653/v1/2021.emnlp-main.699

Wiradinata, A. S., & Mawardi, V. C. (2025). Abstractive Text Summarization Berita Bahasa Indonesia Menggunakan Retrieval-Augmented Generation. *Jurnal Ilmu Komputer Dan Sistem Informasi*, *13*(1). https://doi.org/10.24912/jiksi.v13i1.32861

Wiratmoko, G., Thamrin, H., & Pamungkas, E. W. (2025). Performance of Machine Learning Algorithms on Automatic Summarization of Indonesian Language Texts. *Jurnal Online Informatika*, *10*(1), 196–204. https://doi.org/10.15575/join.v10i1.1506

Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., Barua, A., & Raffel, C. (2021). mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer. *Proceedings of NAACL 2021*, 483–498. https://doi.org/10.18653/v1/2021.naacl-main.41