

IMPLEMENTASI NLP PADA SUMMARIZATION DOKUMEN AKADEMIK MENGUNAKAN TEXTRANK DAN TRANSFORMER MT5

Yaumul Furqan^{1*}, Muhyiddin A.M Hayat², Titin Wahyuni³

^{1,2,3} Universitas Muhammadiyah Makassar, Indonesia

¹105841106822@student.unismuh.ac.id

²titinwahyuni@unismuh.ac.id

³muhyiddin@unismuh.ac.id

* Corresponding Author

Info Artikel

Pengajuan : 31-07-2026
Peninjauan : 13-08-2026
Revisi : 26-08-2026
Diterima : 05-09-2026
Diterbitkan : 16-09-2026

Kata Kunci:

Automatic Text
Summarization, Dokumen
Akademik, ROUGE
Evaluation, TextRank,
Transformer mT5.

Abstrak

Perkembangan dokumen akademik digital meningkatkan kebutuhan akan sistem automatic text summarization yang mampu menyajikan informasi secara ringkas tanpa menghilangkan makna utama. Penelitian ini mengimplementasikan Natural Language Processing (NLP) pada sistem summarization dokumen akademik berbahasa Indonesia berbasis web menggunakan metode TextRank dan Transformer mT5. Dataset terdiri atas dua dokumen akademik yang dievaluasi menggunakan gold summary. Tahapan penelitian meliputi ekstraksi teks, preprocessing, pembangkitan ringkasan, dan evaluasi menggunakan ROUGE-1, ROUGE-2, dan ROUGE-L. Model Transformer mT5 diimplementasikan dengan konfigurasi maksimum input 512 token, output 256 token, beam search 4, length penalty 1,0, dan repetition penalty 1,2. Hasil pengujian menunjukkan bahwa performa kedua metode dipengaruhi oleh karakteristik dokumen. TextRank memperoleh rata-rata ROUGE-1 F1-score sebesar 0,4363, sedangkan Transformer mT5 memperoleh 0,4329. TextRank menghasilkan kesesuaian leksikal yang lebih tinggi terhadap gold summary, sedangkan Transformer mT5 mampu menghasilkan ringkasan yang lebih ringkas berdasarkan hasil observasi. Penelitian ini menunjukkan bahwa kedua pendekatan memiliki karakteristik yang saling melengkapi dalam automatic text summarization dokumen akademik berbahasa Indonesia.

How to Cite: Furqan, Y., Hayat, M. A. M., & Wahyuni, T. (2026). Implementasi NLP pada summarization dokumen akademik menggunakan TextRank dan Transformer mT5. *Journal of Computer Science and Information Technology (JCSIT)*, 3(4), 420–428.

DOI: 10.70248/jcsit.v3i4.4363

This is an open access article under <https://creativecommons.org/licenses/by/4.0/>

Copyright© 2026 by Author. Published by Yayasan Nuraini Ibrahim Mandiri



PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi pada era digital telah mendorong pertumbuhan dokumen akademik, seperti jurnal, skripsi, tesis, disertasi, dan artikel ilmiah, yang kini mudah diakses melalui repositori institusi, perpustakaan digital, maupun platform publikasi elektronik. Kondisi ini memperluas penyebaran ilmu pengetahuan dan akses terhadap informasi ilmiah. Namun, meningkatnya jumlah dokumen juga menimbulkan tantangan dalam proses pencarian, seleksi, dan pemahaman informasi yang relevan secara cepat dan efisien.

Dokumen akademik umumnya memiliki pembahasan yang panjang, kompleks, dan menggunakan bahasa ilmiah yang padat informasi. Mahasiswa, dosen, dan peneliti sering kali harus menelaah banyak referensi untuk memperoleh landasan teori, metodologi, maupun hasil penelitian yang relevan[1]. Proses tersebut membutuhkan waktu yang cukup lama dan berpotensi menurunkan efisiensi karena pengguna tidak selalu dapat mengidentifikasi informasi penting secara cepat.

Permasalahan tersebut menunjukkan perlunya teknologi yang mampu menyajikan inti informasi dokumen secara otomatis tanpa menghilangkan makna utama. Salah satu bidang yang mendukung kebutuhan tersebut adalah *Natural Language Processing* (NLP), yaitu cabang kecerdasan buatan yang memungkinkan komputer memahami, memproses, dan menghasilkan bahasa manusia[2]. Salah satu penerapan NLP yang banyak digunakan adalah *automatic text summarization*, yaitu proses menghasilkan ringkasan secara otomatis dengan mempertahankan informasi penting dan mengurangi redundansi pada dokumen[3].

Secara umum, *text summarization* terdiri atas dua pendekatan, yaitu *extractive summarization* dan *abstractive summarization*. Pendekatan ekstraktif memilih kalimat-kalimat paling representatif dari dokumen asli berdasarkan tingkat relevansinya[4]. Salah satu metode yang banyak digunakan adalah *TextRank*, yaitu algoritma berbasis *graph-based ranking* yang menentukan tingkat kepentingan kalimat melalui hubungan antar kalimat[5]. *TextRank* memiliki keunggulan karena tidak memerlukan data pelatihan, mudah diimplementasikan, dan mampu menghasilkan ringkasan secara efisien dengan biaya komputasi yang relatif rendah.

Sementara itu, perkembangan *deep learning* melahirkan pendekatan abstraktif yang mampu menghasilkan ringkasan dengan bahasa yang lebih natural. Salah satu model yang banyak digunakan adalah *Transformer mT5 (Multilingual Text-to-Text Transfer Transformer)*[6]. Model ini memanfaatkan mekanisme *self-attention* untuk memahami hubungan semantik antarkata dan antarkalimat sehingga mampu menghasilkan ringkasan yang lebih kontekstual, koheren, dan mudah dipahami. Selain mendukung berbagai bahasa, mT5 juga memiliki performa yang baik dalam memproses teks berbahasa Indonesia[7].

Berbagai penelitian mengenai *automatic text summarization* telah dilakukan, namun sebagian besar masih berfokus pada dokumen berbahasa Inggris atau hanya menerapkan satu pendekatan, yaitu ekstraktif atau abstraktif. Pada dokumen akademik berbahasa Indonesia, kajian komparatif kedua pendekatan masih terbatas. Padahal, model berbasis Transformer seperti mT5 berpotensi mengalami *hallucination* dan *information loss* akibat proses parafrase, sedangkan *TextRank* mampu mempertahankan informasi asli, tetapi menghasilkan ringkasan yang kurang natural.[8]. Perbedaan karakteristik tersebut menunjukkan perlunya analisis komparatif untuk mengidentifikasi kelebihan dan keterbatasan masing-masing metode pada proses *summarization* dokumen akademik berbahasa Indonesia.[9].

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan mengimplementasikan dan membandingkan performa metode *TextRank* dan *Transformer mT5* pada sistem *automatic text summarization* dokumen akademik berbahasa Indonesia berbasis web. Sistem yang dikembangkan mencakup proses *preprocessing* teks, pembangkitan ringkasan menggunakan kedua metode, serta evaluasi kualitas ringkasan menggunakan metrik ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*)[10]. Kontribusi penelitian ini meliputi implementasi sistem *summarization* yang mengintegrasikan pendekatan ekstraktif dan abstraktif pada dokumen akademik berbahasa Indonesia, analisis komparatif performa *TextRank* dan *Transformer mT5* berdasarkan hasil evaluasi ROUGE, identifikasi karakteristik masing-masing metode dalam mempertahankan informasi, menghasilkan keterbacaan, dan mendukung proses pemahaman dokumen akademik secara lebih efektif [11].

METODE PENELITIAN

Penelitian ini menggunakan pendekatan pengembangan sistem berbasis

Natural Language Processing (NLP) untuk membangun sistem summarization dokumen akademik berbahasa Indonesia. Objek penelitian berupa dokumen akademik dalam format PDF dan DOCX yang meliputi jurnal, skripsi, tesis, dan artikel penelitian.

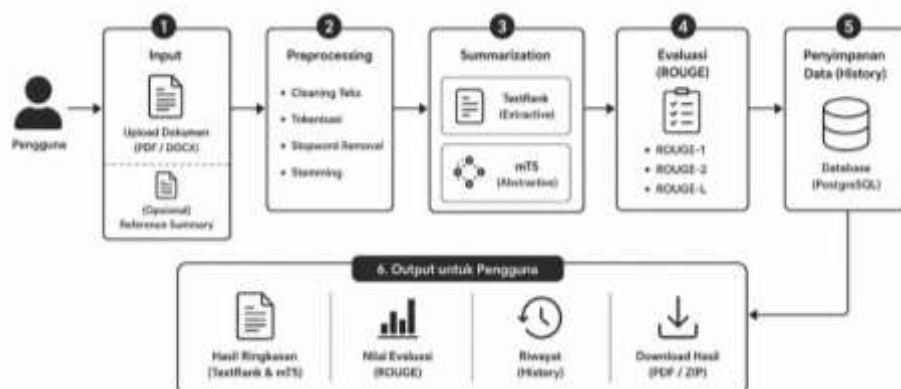
Dataset penelitian terdiri atas dua dokumen akademik berbahasa Indonesia yang digunakan sebagai data uji, masing-masing berupa artikel ilmiah dengan karakteristik panjang dan struktur yang berbeda. Dokumen diperoleh dalam format PDF, kemudian diekstraksi menjadi teks sebelum diproses oleh sistem. Sebagai acuan evaluasi, setiap dokumen dilengkapi dengan *gold summary* yang disusun secara manual berdasarkan informasi utama dokumen sehingga dapat digunakan sebagai referensi pada evaluasi menggunakan metrik ROUGE.

Tahapan penelitian dimulai dari pengumpulan dokumen, kemudian dilanjutkan dengan *preprocessing* teks. Pada metode TextRank, *preprocessing* meliputi *text cleaning*, *case folding*, *tokenisasi*, *stopword removal*, dan *stemming* menggunakan PySastrawi untuk memperoleh representasi kata dasar sebelum pembobotan TF-IDF. Sementara itu, pada Transformer mT5 hanya dilakukan *text cleaning* dan normalisasi dasar tanpa *stemming* karena model memanfaatkan mekanisme *subword tokenizer* yang mempertahankan struktur morfologi kata selama proses tokenisasi. Selanjutnya, sistem menerapkan metode TextRank sebagai pendekatan extractive summarization dengan menghitung keterhubungan antar kalimat menggunakan TF-IDF dan cosine similarity, kemudian memilih kalimat dengan skor tertinggi sebagai ringkasan[12].

Selain itu, penelitian ini mengimplementasikan Transformer mT5 sebagai pendekatan abstractive summarization. Dokumen hasil preprocessing diproses menggunakan model csebuetnlp/mT5_multilingMadeX melalui pustaka HuggingFace Transformers dengan konfigurasi max input length 512 token, max output length 256 token, num beams 4, length penalty 1.0, dan repetition penalty 1.2.

Karena Transformer mT5 memiliki batas maksimum masukan sebesar 512 token, dokumen yang melebihi batas tersebut diproses menggunakan teknik *sentence boundary chunking*. Dokumen dipotong menjadi beberapa *chunk* dengan panjang sekitar 300 kata tanpa memutus struktur kalimat. Setiap *chunk* diringkas secara terpisah, kemudian seluruh hasil ringkasan digabung dan diringkas kembali untuk menghasilkan ringkasan akhir yang lebih koheren.

Kualitas ringkasan dievaluasi menggunakan ROUGE-1, ROUGE-2, dan ROUGE-L berdasarkan nilai *Precision*, *Recall*, dan *F1-score* untuk mengukur tingkat kesamaan informasi antara ringkasan sistem dan *gold summary*. Nilai *overlap* diperoleh dari jumlah unigram, bigram, atau *longest common subsequence* yang sama antara kedua ringkasan sehingga hasil evaluasi dapat menggambarkan tingkat kesesuaian informasi yang dipertahankan oleh masing-masing metode.



Gambar 1. Alur Penelitian Sistem Summarization

Gambar 1 menunjukkan alur penelitian yang dimulai dari ekstraksi dokumen, *preprocessing* teks, pembangkitan ringkasan menggunakan TextRank dan Transformer mT5, kemudian dilanjutkan dengan evaluasi menggunakan metrik ROUGE sebelum hasil ringkasan ditampilkan melalui antarmuka web.

Analisis data dilakukan secara deskriptif komparatif dengan membandingkan performa TextRank dan Transformer mT5 berdasarkan nilai ROUGE (*Precision, Recall*, dan *F1-score*) serta karakteristik hasil ringkasan yang meliputi kesamaan informasi, koherensi kalimat, dan keterbacaan. Analisis ini digunakan untuk mengidentifikasi kelebihan dan keterbatasan masing-masing pendekatan pada proses *automatic text summarization* dokumen akademik berbahasa Indonesia.

Analisis data dilakukan secara deskriptif komparatif dengan membandingkan hasil ringkasan dari metode TextRank dan Transformer mT5 berdasarkan tingkat kesamaan informasi, keterbacaan dan koherensi kalimat, relevansi informasi yang dipertahankan, serta nilai ROUGE F1-score untuk memperoleh gambaran karakteristik penerapan pendekatan *extractive* dan *abstractive summarization* pada dokumen akademik berbahasa Indonesia.

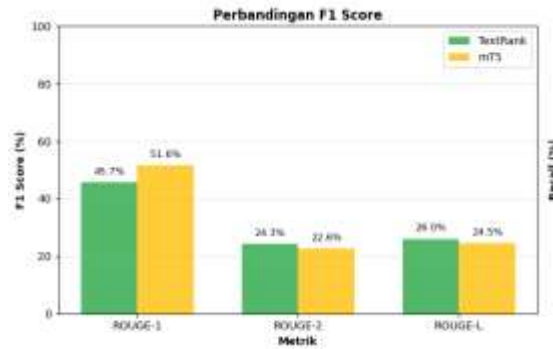
HASIL EVALUASI SUMMARIZATION

Evaluasi kualitas ringkasan dilakukan menggunakan metode ROUGE (Recall-Oriented Understudy for Gisting Evaluation) yang meliputi ROUGE-1, ROUGE-2, dan ROUGE-L dengan fokus pada nilai F1-score. Pada tahap *preprocessing*, sistem melakukan *case folding*, *stopword removal*, dan *stemming* menggunakan PySastrawi untuk mengurangi *noise* dan menyeragamkan bentuk kata. Selanjutnya, TextRank membangun graf berdasarkan bobot TF-IDF dan cosine similarity untuk menentukan kalimat yang paling representatif, sedangkan Transformer mT5 menghasilkan ringkasan secara abstraktif menggunakan arsitektur encoder-decoder dengan mekanisme self-attention serta teknik chunking untuk mengatasi keterbatasan panjang input. Pengujian awal dilakukan menggunakan metrik ROUGE-1 F1-score untuk mengetahui performa metode TextRank dan Transformer mT5 pada setiap dokumen uji. Hasil pengujian disajikan pada Tabel 1.

Tabel 1. Hasil Evaluasi ROUGE-1 F1-score pada Setiap Dokumen Uji

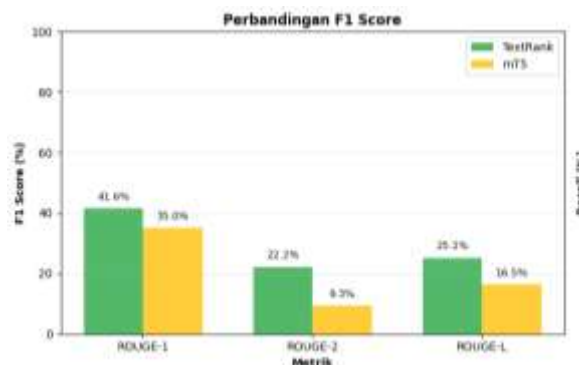
No	Dokumen Uji	TextRank	Transformer mT5
1	ARTIKEL.pdf	0,4566	0,5156
2	It-edu Artikel Ilmiah AI03_Ary_Yulianti_14050974036.pdf	0,4159	0,3502
Rata-Rata		0,4363	0,4329

Berdasarkan Tabel 1, performa kedua metode berbeda pada setiap dokumen. Pada ARTIKEL.pdf, Transformer mT5 memperoleh nilai ROUGE-1 F1-score sebesar 0,5156, lebih tinggi dibandingkan TextRank sebesar 0,4566, dengan selisih 0,0590. Sebaliknya, pada dokumen It-edu Artikel Ilmiah AI03_Ary_Yulianti_14050974036.pdf, TextRank memperoleh nilai 0,4159, lebih tinggi dibandingkan Transformer mT5 yang memperoleh 0,3502, dengan selisih 0,0657. Secara keseluruhan, rata-rata ROUGE-1 F1-score menunjukkan TextRank memperoleh nilai 0,4363, sedikit lebih tinggi dibandingkan Transformer mT5 sebesar 0,4329. Pengujian dilakukan terhadap dua dokumen akademik yang memiliki karakteristik berbeda. Untuk memberikan gambaran performa metode pada setiap dokumen, hasil evaluasi disajikan secara terpisah. Tabel 1 menunjukkan hasil evaluasi pada Dokumen Uji 1, sedangkan Tabel 2 menunjukkan hasil evaluasi pada Dokumen Uji 2.



Gambar 2. Hasil Evaluasi ROUGE pada Dokumen uji 1

Gambar 2 memperjelas hasil pada Tabel 1 bahwa Transformer mT5 memberikan performa terbaik pada dokumen uji 1, sedangkan TextRank lebih unggul pada dokumen kedua. Perbedaan ini menunjukkan bahwa performa kedua metode dipengaruhi oleh karakteristik dokumen yang diringkas. Meskipun demikian, rata-rata nilai ROUGE-1 F1-score kedua metode hanya berbeda 0,0034, sehingga keduanya memiliki performa yang relatif sebanding pada dua dokumen yang diuji. Oleh karena itu, diperlukan evaluasi lanjutan menggunakan metrik ROUGE-1, ROUGE-2, dan ROUGE-L yang mencakup precision, recall, dan F1-score untuk memperoleh gambaran performa yang lebih komprehensif.



Gambar 3. Hasil evaluasi ROUGE pada Dokumen uji 2

Gambar 3 menunjukkan bahwa pada dokumen uji 2 It-edu Artikel Ilmiah AI03_Ary_Yulianti_14050974036.pdf, metode TextRank memperoleh performa lebih baik dibandingkan Transformer mT5. Nilai F1-score ROUGE-1 TextRank mencapai 41,6%, lebih tinggi daripada Transformer mT5 sebesar 35,0%. Pola yang sama juga terlihat pada ROUGE-2 dan ROUGE-L, di mana TextRank secara konsisten memperoleh nilai lebih tinggi. Selain itu, grafik Precision vs Recall menunjukkan bahwa TextRank memiliki keseimbangan precision dan recall yang lebih baik, sehingga menghasilkan F1-score yang lebih tinggi. Berbeda dengan tabel 1, tabel 2 menyajikan hasil evaluasi keseluruhan berdasarkan metrik yang diambil dari jumlah overlap, candidate summary, dan reference summary pada seluruh data pengujian.

HASIL EVALUASI

Hasil evaluasi menunjukkan bahwa performa TextRank dan Transformer mT5 berbeda pada setiap dokumen uji. Pada Dokumen Uji 1, Transformer mT5 memperoleh ROUGE-1 *F1-score* lebih tinggi (0,5156) dibandingkan TextRank (0,4566), sedangkan pada Dokumen Uji 2 TextRank memperoleh nilai lebih tinggi

(0,4159) dibandingkan Transformer mT5 (0,3502). Perbedaan tersebut menunjukkan bahwa performa kedua metode dipengaruhi oleh karakteristik dokumen sehingga evaluasi dilakukan secara independen pada setiap dokumen.

TextRank cenderung menghasilkan nilai ROUGE lebih tinggi karena menggunakan pendekatan *extractive summarization* yang mempertahankan kalimat asli dari dokumen sumber. Karakteristik ini meningkatkan *lexical overlap* terhadap *gold summary*, sehingga nilai *Precision*, *Recall*, dan *F1-score* menjadi lebih tinggi. Sebaliknya, Transformer mT5 menghasilkan ringkasan melalui parafrase sehingga bentuk leksikalnya berbeda dengan *gold summary*. Akibatnya, nilai ROUGE cenderung lebih rendah meskipun informasi utama tetap dipertahankan, dan pada beberapa bagian masih ditemukan *information loss*, terutama pada istilah teknis.

Perlu diperhatikan bahwa ROUGE merupakan metrik berbasis *lexical overlap* yang lebih menguntungkan metode ekstraktif dibandingkan metode abstraktif. Oleh karena itu, evaluasi menggunakan *human evaluation* atau metrik semantik seperti BERTScore dapat dipertimbangkan pada penelitian selanjutnya untuk memberikan penilaian kualitas ringkasan yang lebih komprehensif.

HASIL IMPLEMENTASI WEBSITE

Sistem berhasil mengimplementasikan metode TextRank dan Transformer mT5 pada aplikasi berbasis web yang mendukung proses *automatic text summarization* dokumen akademik. Fitur utama yang disediakan meliputi unggah dokumen, pembangkitan ringkasan menggunakan kedua metode, evaluasi otomatis menggunakan metrik ROUGE, penyimpanan riwayat proses, serta pengunduhan hasil ringkasan.



Gambar 4. Halaman Summarization

Gambar 4 menunjukkan antarmuka proses *summarization*. Pengguna dapat mengunggah dokumen dan *gold summary*, kemudian sistem melakukan ekstraksi teks, *preprocessing*, pembangkitan ringkasan menggunakan TextRank dan Transformer mT5, serta evaluasi otomatis menggunakan ROUGE. Hasil ringkasan dari kedua metode ditampilkan secara berdampingan untuk memudahkan proses analisis.

NO	NAMA DOKUMEN	WAKTU	METODE TERPILIH	Aksi
1	ARTIKEL.pdf	2025-07-11 03:05:25	with Transformer	[Detail] [Download] [Delete]
2	PERKEMBANGAN TEKNOLOGI KECERDASAN BUATAN (AI) DALAM BIDANG KESEHATAN.pdf	2025-07-06 01:56:43	with Transformer	[Detail] [Download] [Delete]
3	PERAN TEKNOLOGI KECERDASAN BUATAN (AI) DALAM BIDANG KESEHATAN.pdf	2025-07-06 03:53:21	with Transformer	[Detail] [Download] [Delete]
4	et al., Jurnal Bina Nusantara, 14(2) 2020: 1-10.pdf	2025-07-05 07:23:11	with Transformer	[Detail] [Download] [Delete]
5	SKRIPSI BOM, metode DMN Falsafah	2025-07-05 07:03:51	with Transformer	[Detail] [Download] [Delete]
6	176-525-1-PR.pdf	2025-07-04 13:47:43	with Transformer	[Detail] [Download] [Delete]
7	ARTIKEL.pdf	2025-07-02 00:13:11	with Transformer	[Detail] [Download] [Delete]

Gambar 5. Halaman Riwayat Summarization

Gambar 5 menampilkan halaman riwayat proses *summarization* yang menyimpan informasi dokumen, waktu pemrosesan, hasil evaluasi, serta fasilitas untuk melihat detail maupun mengunduh hasil ringkasan.

PEMBAHASAN DAN KETERBATASAN

Berdasarkan hasil pengujian, TextRank dan Transformer mT5 menunjukkan karakteristik yang berbeda dalam meringkas dokumen akademik. TextRank memperoleh nilai ROUGE yang lebih tinggi karena mempertahankan kalimat asli sehingga menghasilkan *lexical overlap* yang lebih besar terhadap *gold summary*. Metode ini lebih sesuai digunakan ketika ketepatan informasi faktual menjadi prioritas, seperti pada telaah literatur dan penelusuran referensi ilmiah. Sebaliknya, Transformer mT5 menghasilkan ringkasan yang lebih ringkas, koheren, dan natural melalui proses parafrase. Meskipun nilai ROUGE cenderung lebih rendah akibat berkurangnya kesamaan leksikal, metode ini lebih sesuai untuk memberikan gambaran umum dokumen secara cepat dan mudah dipahami. Hasil penelitian menunjukkan bahwa pemilihan metode perlu disesuaikan dengan kebutuhan pengguna. TextRank lebih tepat untuk menjaga akurasi informasi, sedangkan Transformer mT5 lebih unggul dalam keterbacaan. Kombinasi kedua pendekatan berpotensi menjadi solusi yang lebih efektif untuk pengembangan sistem *automatic text summarization* pada penelitian selanjutnya.

KESIMPULAN

Berdasarkan hasil penelitian, kedua metode memiliki karakteristik yang berbeda dalam menghasilkan ringkasan dokumen akademik. TextRank lebih sesuai digunakan pada dokumen yang membutuhkan ketepatan informasi dan kesesuaian terhadap dokumen sumber, sedangkan Transformer mT5 lebih sesuai digunakan apabila keterbacaan dan kealamian bahasa menjadi prioritas. Penelitian ini masih memiliki keterbatasan karena hanya menggunakan dua dokumen uji dan evaluasi dilakukan menggunakan metrik ROUGE tanpa melibatkan penilaian manusia. Penelitian selanjutnya dapat menggunakan dataset yang lebih besar, melakukan *fine-tuning* model mT5 pada korpus akademik berbahasa Indonesia, serta menambahkan evaluasi semantik maupun *human evaluation* untuk memperoleh gambaran kualitas ringkasan yang lebih komprehensif.

Pernyataan Mengenai Penggunaan Kecerdasan Buatan (AI)

Dalam penyusunan naskah ini, penulis menggunakan perangkat kecerdasan buatan (AI) sebagai alat bantu dalam memperbaiki tata bahasa, menyusun kalimat,

dan meningkatkan kejelasan penyampaian. Penggunaan AI tidak menggantikan proses penelitian, analisis, maupun tanggung jawab penulis terhadap isi naskah. Seluruh bagian yang dibantu oleh AI telah diperiksa dan disesuaikan oleh penulis agar tetap sesuai dengan hasil penelitian. Penulis bertanggung jawab atas keakuratan, orisinalitas, dan integritas naskah yang disampaikan.

Konflik Kepentingan

Penulis menyatakan bahwa tidak terdapat konflik kepentingan yang berkaitan dengan penelitian maupun publikasi artikel ini.

Kontribusi Penulis

Yaumul Furqan berperan dalam perancangan penelitian, pengembangan sistem, pengolahan data, implementasi metode, analisis hasil, dan penyusunan naskah. Muhyiddin A.M Hayat dan Titin Wahyuni berperan dalam memberikan arahan penelitian, melakukan peninjauan terhadap hasil, serta memberikan masukan dalam penyempurnaan naskah. Seluruh penulis telah membaca dan menyetujui versi akhir artikel.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Universitas Muhammadiyah Makassar, khususnya Fakultas Teknik dan Program Studi Informatika, atas dukungan yang diberikan selama pelaksanaan penelitian. Penulis juga menyampaikan terima kasih kepada dosen pembimbing atas arahan, masukan, dan bimbingan selama proses penelitian dan penyusunan naskah ini. Ucapan terima kasih turut disampaikan kepada semua pihak yang telah membantu sehingga penelitian ini dapat diselesaikan.

DAFTAR PUSTAKA

- [1] G. Sharma, D. Sharma, and M. Sasikumar, "Summarizing long scientific documents through hierarchical structure extraction," *Nat. Lang. Process. J.*, vol. 8, no. December 2023, p. 100080, 2024, doi: 10.1016/j.nlp.2024.100080.
- [2] H. Wang, "Automatic question-answering modeling in English by integrating TF-IDF and segmentation algorithms," *Syst. Soft Comput.*, vol. 6, no. February, p. 200087, 2024, doi: 10.1016/j.sasc.2024.200087.
- [3] S. Goyal and A. Kaur, "Summarization of Software Bug Report based on Sentence Semantic Similarity (SSBRSSS) Technique," *Procedia Comput. Sci.*, vol. 235, no. 2023, pp. 1761–1771, 2024, doi: 10.1016/j.procs.2024.04.167.
- [4] S. S. Abdulkareem, "Hybrid extractive – abstractive summarization of scientific texts via deep clustering and attention mechanism," vol. 138, no. December 2025, pp. 21–35, 2026.
- [5] M. F. Fakhrezi, M. A. Bijaksana, and A. F. Huda, "Implementation of Automatic Text Summarization with TextRank Method in the Development of Al-Qur'an Vocabulary Encyclopedia," *Procedia Comput. Sci.*, vol. 179, no. 2020, pp. 391–398, 2021, doi: 10.1016/j.procs.2021.01.021.
- [6] S. Phani, A. Abdul, M. K. S. Prasad, and V. D. Reddy, "MATSTF: User query-based multilingual abstractive text summarization for low resource Indian languages by fine-tuning mT5," *Alexandria Eng. J.*, vol. 127, no. December 2024, pp. 129–142, 2025, doi: 10.1016/j.aej.2025.04.031.
- [7] N. M. Gardazi, M. K. Malik, and A. Daud, "Pahari POS-tagged corpus: A large-scale linguistic resource for NLP applications," *Data Br.*, vol. 65, 2026, doi: 10.1016/j.dib.2026.112515.

- [8] A. Auriemma Citarella, M. Barbella, M. G. Ciobanu, F. De Marco, L. Di Biasi, and G. Tortora, "Assessing the effectiveness of ROUGE as unbiased metric in Extractive vs. Abstractive summarization techniques," *J. Comput. Sci.*, vol. 87, no. February 2024, p. 102571, 2025, doi: 10.1016/j.jocs.2025.102571.
- [9] S. Yunus, C. Hark, and F. Okumuş, "Enhancing extractive multi-documents summarization with a novel dominating set model for semantic relationship detection," *Eng. Sci. Technol. an Int. J.*, vol. 69, no. July, 2025, doi: 10.1016/j.jestch.2025.102127.
- [10] A. Goswami, "ScienceDirect Blockchain-Integrated Network for Secure Data Sharing in Collaborative NLP Research Applications," *Procedia Comput. Sci.*, vol. 275, pp. 200–207, 2026, doi: 10.1016/j.procs.2026.01.025.
- [11] D. Zare, L. Fernandez-Sanz, V. Pospelova, I. Lopez-Balominos, N. Rizun, and A. Revina, "Semantic mapping of soft skills from skillsmatch framework to the standard EN 16234-1 using transformer-based NLP," *Comput. Stand. Interfaces*, vol. 98, no. March, 2026, doi: 10.1016/j.csi.2026.104162.
- [12] J. S. Zhu, H. Sun, and B. Kong, "Improving multilingual English translation performance through T5 and MAML integration," *Syst. Soft Comput.*, vol. 7, no. October 2024, p. 200394, 2025, doi: 10.1016/j.sasc.2025.200394.