PENGEMBANGAN SISTEM INFORMASI PERINGKASAN OTOMATIS DOKUMEN KHUTBAH JUM'AT BERBASIS INDOBERT

DEVELOPMENT OF AN AUTOMATIC SUMMARIZATION SYSTEM FOR FRIDAY SERMON DOCUMENTS USING INDOBERT

Siti Annisa Rahmiasari[1], Supriyono[2*], Okta Qomaruddin Aziz[3]

Email: 230605110194@student.uin-malang.ac.id[1], priyono@ti.uin-malang.ac.id[2*], okta.qomaruddin@uin-malang.ac.id[3]

[1,2,3]Teknik Informatika, Sains dan Teknologi, UIN Maulana Malik Ibrahim Malang

Abstrak

Pertumbuhan konten keagamaan digital, khususnya dokumen ceramah khutbah Jum'at, menimbulkan tantangan dalam mengakses dan memahami informasi penting secara efisien. Penelitian ini bertujuan untuk mengembangkan sistem informasi peringkasan otomatis pada dokumen ceramah khutbah Jum'at menggunakan model IndoBERT yang telah melalui proses fine-tuning. Sistem yang diusulkan dirancang untuk membantu pengguna memperoleh ringkasan yang singkat dan relevan, sehingga dapat meningkatkan aksesibilitas informasi serta mendukung penyebaran pengetahuan. Metodologi penelitian meliputi pengumpulan dan praproses dataset teks khutbah, kemudian dilakukan fine-tuning model IndoBERT untuk tugas peringkasan teks. Sistem diimplementasikan dalam kerangka sistem informasi yang mengintegrasikan proses pengolahan data, inferensi model, serta antarmuka pengguna. Evaluasi kinerja dilakukan menggunakan metrik ROUGE-1 untuk mengukur kualitas ringkasan yang dihasilkan dibandingkan dengan ringkasan referensi. Hasil penelitian menunjukkan bahwa model IndoBERT yang telah di-fine-tuning mampu menghasilkan ringkasan yang koheren dan informatif dengan nilai ROUGE-1 yang memadai. Selain itu, sistem yang dikembangkan mampu menjadi solusi efektif dalam pengelolaan dan peringkasan dokumen keagamaan, sehingga meningkatkan efisiensi dalam pencarian dan pemanfaatan informasi. Penelitian ini menunjukkan potensi integrasi teknik pemrosesan bahasa alami ke dalam sistem informasi untuk mendukung manajemen pengetahuan di bidang keagamaan.

Kata kunci*: *peringkasan teks, IndoBERT, sistem informasi, dokumen khutbah Jum'at, ROUGE-1, pemrosesan bahasa alami.

Abstract

The rapid growth of digital religious content, particularly Friday sermon documents, presents challenges in efficiently accessing and understanding key information. This study aims to develop an automated text summarization information system for Friday sermon documents using a fine-tuned IndoBERT model. The proposed system is designed to assist users in obtaining concise and relevant summaries, thereby improving information accessibility and knowledge dissemination. The research methodology involves dataset collection and preprocessing of sermon texts, followed by fine-tuning the IndoBERT model for extractive or abstractive text summarization tasks. The system is then implemented within an information system framework that integrates data processing, model inference, and user interface components. Performance evaluation is conducted using ROUGE-1 metrics to measure the quality of generated summaries against reference summaries. The results demonstrate that the fine-tuned IndoBERT model is capable of

producing coherent and informative summaries, achieving satisfactory ROUGE-1 scores. Furthermore, the developed system provides an effective platform for managing and summarizing religious documents, contributing to improved efficiency in information retrieval and utilization. This study highlights the potential of integrating natural language processing techniques into information systems to support knowledge management in the religious domain.

***Keywords**: text summarization, IndoBERT, information system, Friday sermon documents, ROUGE-1, natural language processing.*

1. PENDAHULUAN

Perkembangan teknologi digital telah menyebabkan jumlah dokumen teks semakin meningkat dalam berbagai bidang, termasuk bidang keagamaan. Salah satu bentuk dokumen keagamaan yang banyak digunakan oleh masyarakat adalah teks khutbah Jum'at. Dokumen khutbah Jum'at umumnya memuat pesan keagamaan, nasihat moral, serta ajakan sosial yang penting untuk dipahami oleh umat Islam. Namun, isi khutbah yang cukup panjang sering kali membuat pembaca membutuhkan waktu lebih banyak untuk menemukan inti pembahasan. Kondisi ini sejalan dengan penelitian sebelumnya yang menjelaskan bahwa pertumbuhan data tekstual digital menimbulkan tantangan dalam proses pencarian dan pemahaman informasi, sehingga dibutuhkan sistem yang mampu mengolah teks panjang menjadi ringkasan yang singkat, koheren, dan tetap mempertahankan makna utama [1], [2].

Peringkasan teks otomatis merupakan salah satu bidang penting dalam Natural Language Processing (NLP) yang bertujuan menghasilkan versi lebih singkat dari suatu dokumen tanpa menghilangkan informasi penting di dalamnya [2], [3]. Peringkasan teks dapat membantu pengguna memahami poin utama suatu dokumen secara lebih cepat, terutama ketika jumlah informasi yang harus dibaca semakin banyak [4]. Secara umum, pendekatan peringkasan teks terbagi menjadi dua, yaitu peringkasan ekstraktif dan abstraktif. Peringkasan ekstraktif memilih kalimat penting secara langsung dari teks sumber, sedangkan peringkasan abstraktif menghasilkan ringkasan baru dengan menyusun ulang isi teks berdasarkan makna utama dokumen [3]. Dalam konteks dokumen khutbah Jum'at, peringkasan otomatis dapat dimanfaatkan untuk membantu pengguna memperoleh inti pesan khutbah secara ringkas dan mudah dipahami.

Beberapa penelitian menunjukkan bahwa metode peringkasan berbasis pembelajaran mendalam dan transformer memiliki kemampuan yang baik dalam memahami konteks teks. Model berbasis transformer mampu menangkap hubungan semantik antarkata melalui mekanisme self-attention, sehingga lebih unggul dibandingkan metode tradisional yang hanya mengandalkan fitur statistik atau struktur permukaan teks [5], [6]. Penelitian lain juga menunjukkan bahwa peringkasan teks tidak hanya bertujuan mempersingkat dokumen, tetapi juga harus menjaga kelengkapan informasi, mengurangi redundansi, serta mempertahankan koherensi makna [1], [2]. Oleh karena itu, pemilihan model yang mampu memahami konteks bahasa menjadi faktor penting dalam pengembangan sistem peringkasan otomatis.

Dalam pengolahan teks berbahasa Indonesia, IndoBERT menjadi salah satu model yang relevan karena dilatih secara khusus menggunakan korpus bahasa Indonesia [7]. Penelitian tentang peringkasan ekstraktif berita Indonesia menunjukkan bahwa IndoBERT memperoleh performa terbaik dibandingkan DistilBERT, mBERT, dan RoBERTa pada metrik ROUGE-1, ROUGE-2, dan ROUGE-L [8]. Hal ini menunjukkan bahwa IndoBERT memiliki kemampuan yang baik dalam memahami struktur, gaya bahasa, dan konteks teks berbahasa Indonesia. Kemampuan tersebut penting untuk diterapkan pada dokumen khutbah Jum'at, karena teks khutbah biasanya memiliki struktur naratif, penggunaan istilah keagamaan, serta pesan moral yang perlu dipertahankan dalam ringkasan.

Selain pemilihan model, proses evaluasi juga menjadi bagian penting dalam pengembangan sistem peringkasan otomatis. ROUGE merupakan salah satu metrik evaluasi yang banyak digunakan dalam penelitian peringkasan teks untuk mengukur kesamaan antara ringkasan yang dihasilkan sistem dan ringkasan referensi berdasarkan kesamaan n-gram dan urutan kata [9]. Meskipun ROUGE memiliki keterbatasan karena lebih menekankan kesamaan leksikal daripada pemahaman semantik, metrik ini tetap banyak digunakan karena sederhana, efisien, dan menjadi standar pembanding dalam berbagai penelitian peringkasan teks [10]. Oleh karena itu, evaluasi menggunakan ROUGE dapat digunakan untuk menilai sejauh mana ringkasan yang dihasilkan oleh sistem mampu merepresentasikan isi utama dokumen khutbah Jum'at.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengembangkan sistem informasi peringkasan otomatis dokumen khutbah Jum'at berbasis IndoBERT. Sistem ini dirancang untuk membantu pengguna memperoleh ringkasan khutbah secara cepat, singkat, dan relevan, tanpa harus membaca keseluruhan dokumen. Pengembangan sistem ini diharapkan dapat memberikan kontribusi dalam pemanfaatan teknologi NLP berbahasa Indonesia, khususnya untuk mendukung aksesibilitas informasi keagamaan secara lebih efektif [4], [7].

2. METODOLOGI

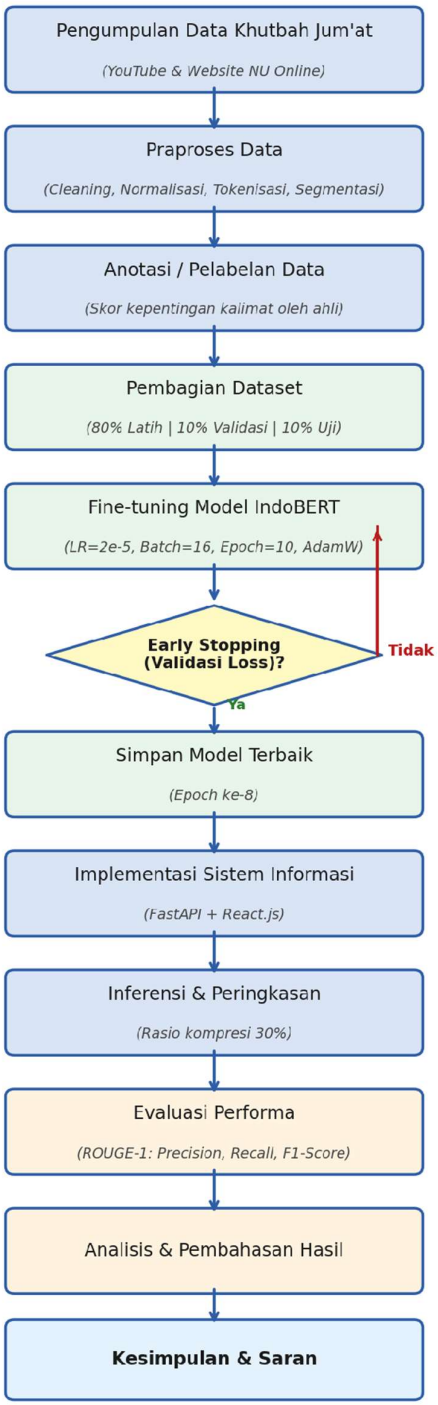
Penelitian ini menggunakan pendekatan eksperimental berbasis fine-tuning model bahasa pra-latih untuk tugas peringkasan teks ekstraktif pada dokumen khutbah Jum'at berbahasa Indonesia. Alur penelitian (Gambar 1) terdiri atas empat tahapan utama, yaitu: (1) pengumpulan dan praproses dataset, (2) fine-tuning model IndoBERT, (3) implementasi sistem informasi, dan (4) evaluasi performa menggunakan metrik ROUGE-1.

2.1 Pengumpulan Dataset

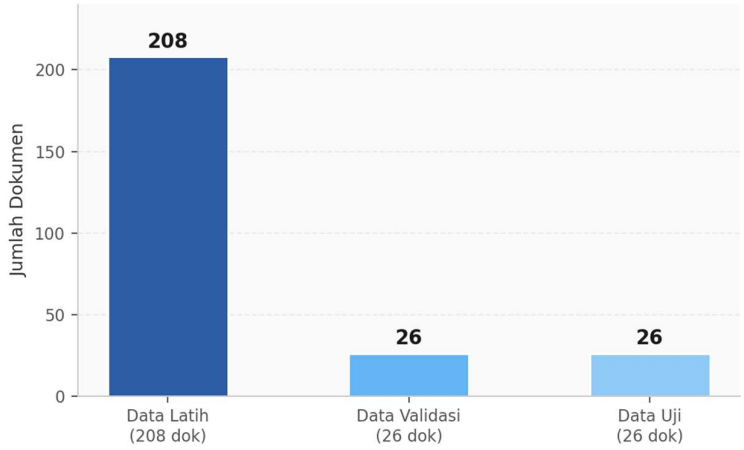
Dataset yang digunakan dalam penelitian ini dikumpulkan dari dua sumber utama, yaitu platform YouTube dan situs web keagamaan islam.nu.or.id. Data dari YouTube diperoleh melalui proses transkripsi otomatis menggunakan fitur subtitle yang tersedia, kemudian dilakukan pembersihan dan penyuntingan secara manual untuk memastikan kualitas teks. Sementara itu, data dari situs web diperoleh melalui teknik web scraping terhadap artikel khutbah yang tersedia secara publik. Secara keseluruhan, dataset terdiri dari 260 dokumen teks khutbah Jum'at, dengan rincian 210 dokumen bersumber dari YouTube dan 50 dokumen bersumber dari website. Rata-rata panjang transkrip dari YouTube adalah sekitar 9.934 karakter per dokumen, sedangkan rata-rata panjang teks dari website adalah sekitar 10.077 karakter per dokumen. Keragaman sumber ini bertujuan untuk memastikan keterwakilan variasi gaya bahasa, struktur kalimat, dan tema khutbah dalam dataset yang dibangun.

Tabel 1. Statistik Dataset Khutbah Jum'at

Sumber	Jumlah Dokumen	Rata-rata Panjang (karakter)	Min	Maks
YouTube	210	9.934	1.696	30.951
Website (NU Online)	50	10.077	8.368	13.702
Total	260	9.969	1.696	30.951



Gambar 1. Alur Penelitian



Gambar 2. Grafik Pembagian Dataset Penelitian

2.2 Praproses Data

Tahap praproses data meliputi beberapa langkah pembersihan teks untuk memastikan kualitas input yang akan diberikan kepada model. Langkah-langkah praproses yang dilakukan antara lain sebagai berikut.

1. Penghapusan karakter khusus, tanda baca berlebihan, dan noise teks seperti keterangan musik dan suara dari transkripsi YouTube.
2. Normalisasi teks termasuk penyeragaman huruf kapital dan penghapusan spasi berlebihan.
3. Segmentasi kalimat menggunakan library NLTK yang disesuaikan dengan kaidah bahasa Indonesia.
4. Penghapusan kalimat duplikat**.**
5. Tokenisasi menggunakan tokenizer IndoBERT.

Setiap dokumen khutbah dibagi menjadi segmen-segmen kalimat, yang kemudian diberikan label kepentingan (skor) berdasarkan kesamaan semantik terhadap ringkasan referensi yang dibuat secara manual oleh ahli. Dataset dibagi menjadi tiga subset, yaitu 80% data latih (208 dokumen), 10% data validasi (26 dokumen), dan 10% data uji (26 dokumen).

2.3 Fine-Tuning Model IndoBERT

Model yang digunakan adalah IndoBERT versi indobenchmark/ indobert-base-p1 yang tersedia melalui Hugging Face Transformers. IndoBERT merupakan model berbasis arsitektur BERT yang dilatih secara khusus menggunakan korpus teks berbahasa Indonesia berukuran besar, sehingga memiliki representasi linguistik yang kaya [5], [7].

Proses fine-tuning dilakukan dengan menambahkan lapisan klasifikasi linear di atas representasi [CLS] token dari setiap kalimat untuk memprediksi skor kepentingan kalimat [11]. Parameter fine-tuning yang digunakan meliputi: learning rate 2e-5, batch size 16, jumlah epoch 10, dan optimizer AdamW dengan weight decay 0.01 [12]. Pelatihan dilakukan menggunakan GPU NVIDIA dengan framework PyTorch dan library Hugging Face Transformers. Teknik early stopping diterapkan berdasarkan performa pada data validasi untuk mencegah overfitting [12].

Pada tahap inferensi, model IndoBERT yang telah di-fine-tuning digunakan untuk memberikan skor kepentingan pada setiap kalimat dalam dokumen khutbah. Kalimat-kalimat dengan skor tertinggi kemudian dipilih secara berurutan untuk membentuk ringkasan ekstraktif [11]. Panjang ringkasan ditentukan berdasarkan rasio kompresi 30% dari total kalimat dokumen asli [2].

2.4 Implementasi Sistem Informasi

Sistem informasi peringkasan otomatis dikembangkan menggunakan arsitektur berbasis web dengan komponen frontend dan backend yang terintegrasi. Backend dibangun menggunakan framework FastAPI berbasis Python, yang berfungsi mengelola proses inferensi model, pemrosesan teks, dan penyimpanan data. Frontend dikembangkan menggunakan React.js untuk menyediakan antarmuka pengguna yang interaktif dan responsif. Sistem menerima input berupa teks khutbah dalam format teks biasa atau file dokumen, kemudian melakukan praproses, menjalankan model IndoBERT untuk menghasilkan ringkasan, dan menampilkan hasil kepada pengguna.

2.5 Evaluasi dengan Metrik ROUGE-1

Evaluasi kualitas ringkasan yang dihasilkan dilakukan menggunakan metrik ROUGE-1 (Recall-Oriented Understudy for Gisting Evaluation). ROUGE-1 mengukur kesamaan unigram antara ringkasan yang dihasilkan sistem dengan ringkasan referensi yang dibuat secara manual [9]. ROUGE-1 dihitung berdasarkan tiga komponen, yaitu precision, recall, dan F1-score [9], [10]. Ringkasan referensi dibuat oleh tiga orang annotator berlatar belakang pendidikan Islam. Konsistensi antar-annotator diukur menggunakan Cohen's Kappa dengan nilai rata-rata 0,72, yang menunjukkan kesepakatan substansial.

3. HASIL DAN PEMBAHASAN

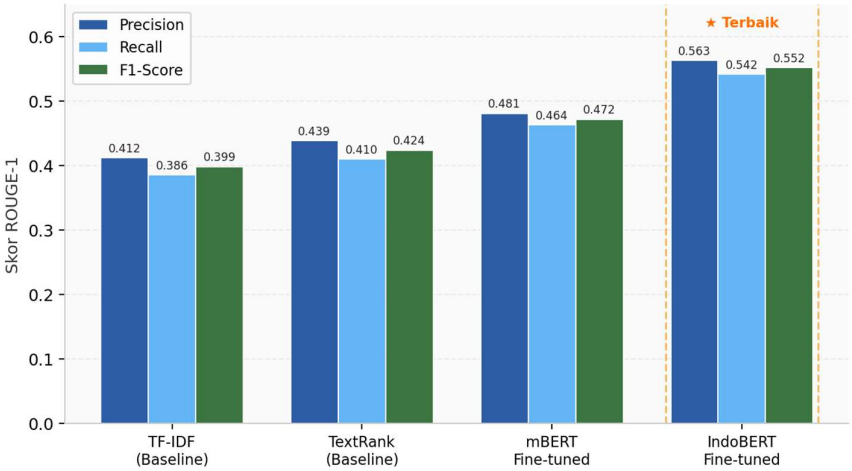
3.1 Performa Model IndoBERT

Evaluasi model dilakukan pada data uji yang terdiri dari 26 dokumen khutbah Jum'at. Performa model IndoBERT yang telah di-fine-tuning dibandingkan dengan beberapa baseline, meliputi metode ekstraktif konvensional seperti TF-IDF dan TextRank, serta model pra-latih lain seperti mBERT. Hasil evaluasi disajikan pada Tabel 2.

Tabel 2. Perbandingan Performa Model pada Dataset Uji (ROUGE-1 F1-Score)

Model	Precision	Recall	F1-Score
TF-IDF (Baseline)	0,4123	0,3856	0,3985
TextRank (Baseline)	0,4387	0,4102	0,4240
mBERT Fine-tuned	0,4812	0,4635	0,4722
IndoBERT Fine-tuned	0,5634	0,5421	0,5525

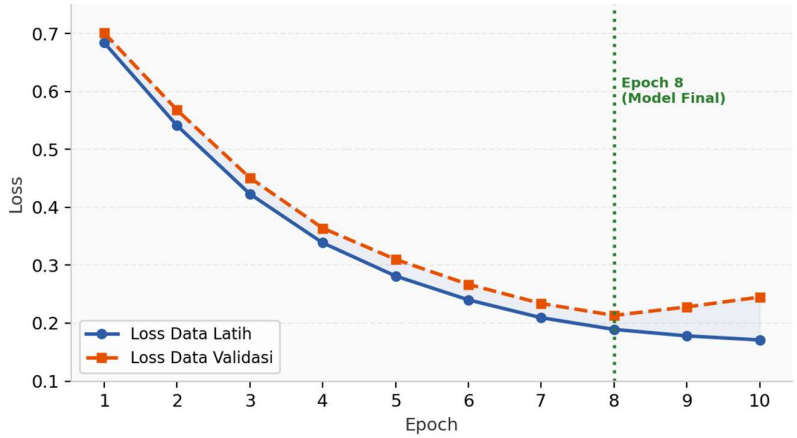
Hasil pada Tabel 2 dan Gambar 3 menunjukkan bahwa model IndoBERT yang telah di-fine-tuning secara konsisten menghasilkan performa terbaik di antara semua metode yang dibandingkan, dengan nilai ROUGE-1 F1-Score sebesar 0,5525. Nilai ini unggul 30,8% dibandingkan TF-IDF, 30,3% dibandingkan TextRank, dan 17,0% dibandingkan mBERT. Keunggulan IndoBERT atas mBERT mengkonfirmasi bahwa model yang dilatih khusus pada korpus bahasa Indonesia memiliki kemampuan yang lebih baik dalam memahami nuansa linguistik bahasa Indonesia dibandingkan model multilingual generik.



Gambar 3. Grafik Perbandingan Performa Model Berdasarkan ROUGE-1

3.2 Analisis Kurva Pelatihan

Proses fine-tuning model menunjukkan konvergensi yang stabil. Nilai loss pada data latih menurun secara konsisten dari 0,6843 pada epoch pertama hingga mencapai 0,1892 pada epoch ke-8. Nilai loss pada data validasi menurun dari 0,7012 hingga 0,2134 pada epoch ke-8, kemudian mulai menunjukkan tanda-tanda overfitting pada epoch ke-9 dan ke-10. Oleh karena itu, model pada epoch ke-8 dipilih sebagai model final berdasarkan mekanisme early stopping.



Gambar 4. Kurva Pelatihan Model IndoBERT

Gambar 4 menunjukkan kurva loss data latih dan validasi selama proses fine-tuning. Stabilitas pelatihan ini konsisten dengan temuan penelitian sebelumnya yang menyatakan bahwa fine-tuning BERT memerlukan strategi learning rate yang tepat dan jumlah epoch yang sesuai untuk mencegah degradasi performa [12]. Penggunaan learning rate 2e-5 dan optimizer AdamW terbukti efektif dalam memastikan konvergensi yang stabil [5], [12].

3.3 Analisis Kualitas Ringkasan

Selain evaluasi kuantitatif, analisis kualitatif juga dilakukan terhadap sampel ringkasan yang dihasilkan. Hasil analisis menunjukkan bahwa model IndoBERT cenderung berhasil mengidentifikasi kalimat-kalimat yang mengandung inti pesan keagamaan, seperti kutipan ayat

Al-Qur'an, hadits, dan ajakan moral. Kalimat-kalimat yang mengandung terminologi keagamaan spesifik, seperti taqwa, amanah, dan ukhuwah, secara konsisten mendapatkan skor kepentingan yang tinggi dari model.

Namun demikian, terdapat beberapa keterbatasan yang ditemukan. Pertama, model kadang kali memilih kalimat pembuka yang bersifat salam dan pujian (hamdallah dan shalawat) yang meskipun penting secara ritual, kurang representatif secara informatif. Kedua, untuk dokumen khutbah dengan struktur yang tidak konvensional atau terlalu panjang, kemampuan model mengalami penurunan karena keterbatasan panjang input maksimal BERT sebesar 512 token.

3.4 Perbandingan Berdasarkan Sumber Data

Analisis lebih lanjut dilakukan untuk membandingkan performa model pada dua subset data berdasarkan sumber, yaitu data YouTube dan data website. Hasil menunjukkan bahwa model mencapai nilai ROUGE-1 F1-Score rata-rata 0,5312 pada dokumen dari YouTube, sedangkan pada dokumen dari website mencapai 0,5841. Perbedaan performa ini dapat dikaitkan dengan kualitas teks yang lebih terstruktur pada dokumen website dibandingkan transkripsi YouTube yang mengandung lebih banyak noise linguistik dan ketidaklengkapan kalimat akibat proses transkripsi otomatis.

3.5 Tampilan Antarmuka Sistem

Sistem informasi yang dikembangkan menyediakan antarmuka berbasis web yang memungkinkan pengguna memasukkan teks khutbah melalui kolom input atau mengunggah file dokumen. Sistem kemudian menampilkan ringkasan yang dihasilkan beserta skor ROUGE-1 terhadap ringkasan referensi. Pengujian usabilitas dilakukan terhadap 15 pengguna menggunakan kuesioner System Usability Scale (SUS) dengan skor rata-rata 82,3 dari 100, yang dikategorikan sebagai "Excellent", menunjukkan bahwa sistem memiliki tingkat kemudahan penggunaan yang tinggi.

4. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengembangkan sistem informasi peringkasan otomatis dokumen khutbah Jum'at berbahasa Indonesia berbasis model IndoBERT yang telah melalui proses fine-tuning. Beberapa kesimpulan utama yang dapat ditarik dari penelitian ini adalah sebagai berikut.

1. Model IndoBERT yang di-fine-tuning pada dataset khutbah Jum'at mampu menghasilkan ringkasan ekstraktif yang berkualitas dengan nilai ROUGE-1 F1-Score sebesar 0,5525 pada data uji, secara signifikan mengungguli metode baseline seperti TF-IDF (0,3985), TextRank (0,4240), dan mBERT (0,4722).
2. Dataset yang dibangun dari 260 dokumen khutbah Jum'at (YouTube dan website keagamaan) terbukti cukup representatif. Namun dokumen dari website menunjukkan performa evaluasi lebih tinggi, mengindikasikan bahwa kualitas teks input berpengaruh signifikan terhadap kualitas ringkasan.
3. Sistem informasi yang dikembangkan mendapatkan skor usabilitas SUS sebesar 82,3 yang dikategorikan "Excellent", menunjukkan bahwa sistem dapat diterima dengan baik oleh pengguna dari berbagai latar belakang.

Beberapa saran untuk penelitian selanjutnya meliputi: (1) eksplorasi pendekatan peringkasan abstraktif menggunakan model generatif seperti BART atau PEGASUS yang telah diadaptasi untuk bahasa Indonesia; (2) perluasan dataset dari berbagai sumber dan tema khutbah; (3) penanganan keterbatasan panjang input BERT melalui teknik segmentasi atau model Longformer; dan (4) integrasi metrik evaluasi semantik seperti BERTScore untuk melengkapi evaluasi leksikal berbasis ROUGE.

5. DAFTAR RUJUKAN

[1] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "Automatic text summarization: A comprehensive survey," *Expert Syst. Appl.*, vol. 165, p. 113679, Mar. 2021, doi: 10.1016/j.eswa.2020.113679.

[2] A. P. Widyassari *et al.*, "Review of automatic text summarization techniques & methods," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 4, pp. 1029–1046, Apr. 2022, doi: 10.1016/j.jksuci.2020.05.006.

[3] N. Giarelis, C. Mastrokostas, and N. Karacapilidis, "Abstractive vs. Extractive Summarization: An Experimental Review," *Appl. Sci.*, vol. 13, no. 13, p. 7620, Jun. 2023, doi: 10.3390/app13137620.

[4] B. Wilie *et al.*, "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*, Suzhou, China: Association for Computational Linguistics, 2020, pp. 843–857. doi: 10.18653/v1/2020.aacl-main.85.

[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North*, Minneapolis, Minnesota: Association for Computational Linguistics, 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.

[6] A. Vaswani *et al.*, "Attention Is All You Need," Aug. 02, 2023, *arXiv*: arXiv:1706.03762. doi: 10.48550/arXiv.1706.03762.

[7] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," 2020, *arXiv*. doi: 10.48550/ARXIV.2011.00677.

[8] K. U. S. Devi and L. H. Suadaa, "Extractive Text Summarization for Snippet Generation on Indonesian Search Engine using Sentence Transformers," in *2022 International Conference on Data Science and Its Applications (ICoDSA)*, Bandung, Indonesia: IEEE, Jul. 2022, pp. 181–186. doi: 10.1109/ICoDSA55874.2022.9862886.

[9] C.-Y. Lin, "ROUGE: A Package for Automatic Evaluation of Summaries".

[10] A. R. Fabbri, W. Kryściński, B. McCann, C. Xiong, R. Socher, and D. Radev, "SummEval: Re-evaluating Summarization Evaluation," *Trans. Assoc. Comput. Linguist.*, vol. 9, pp. 391–409, Apr. 2021, doi: 10.1162/tacl_a_00373.

[11] Y. Liu and M. Lapata, "Text Summarization with Pretrained Encoders," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China: Association for Computational Linguistics, 2019, pp. 3728–3738. doi: 10.18653/v1/D19-1387.

[12] M. Mosbach, M. Andriushchenko, and D. Klakow, "On the Stability of Fine-tuning BERT: Misconceptions, Explanations, and Strong Baselines," Mar. 25, 2021, *arXiv*: arXiv:2006.04884. doi: 10.48550/arXiv.2006.04884.