



Pengembangan Model BERT dengan Metode *Abstraction Summarization* untuk Ringkasan Otomatis Tafsir Ayat Al-Qur'an

Proposal Skripsi

diajukan sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer

oleh
Reiki Aziz Yoga Utama
4611422055

**TEKNIK INFORMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI SEMARANG
SEMARANG
2024**

PERSETUJUAN PEMBIMBING

Skripsi berjudul “Pengembangan Model BERT dengan Metode *Abstraction Summarization* untuk Ringkasan Otomatis Tafsir Ayat Al-Qur'an” yang disusun oleh :

nama	: Reiki Aziz Yoga Utama
NIM	: 4611422055
Prodi/Fakultas	: Teknik Informatika/Matematika dan Ilmu Pengetahuan Alam

Semarang, 06 September 2024
Belum ditentukan

Belum ditentukan & NIP

BAB I

PENDAHULUAN

1.1 Latar Belakang

Al-Qur'an merupakan kitab suci umat Islam yang menjadi sumber hukum Islam yang pertama dan utama (Aji Fitra Jaya Institut Perguruan Tinggi Ilmu Al Qur, n.d.). Dalam praktiknya, pemahaman terhadap Al-Qur'an seringkali membutuhkan penjelasan atau tafsir yang mendalam agar pesan yang terkandung dapat dimengerti oleh berbagai kalangan. Tafsir Al-Qur'an adalah karya ilmiah yang menjelaskan makna ayat-ayat Al-Qur'an secara mendetail, baik dalam bahasa Arab maupun bahasa lainnya, termasuk bahasa Indonesia. Namun, panjangnya tafsir sering kali menjadi tantangan, terutama bagi mereka yang membutuhkan pemahaman cepat namun tetap akurat.

Seiring dengan perkembangan teknologi, berbagai pendekatan berbasis kecerdasan buatan (Artificial Intelligence) mulai diterapkan untuk mempermudah akses terhadap informasi, termasuk dalam bidang teks keagamaan. Salah satu teknologi yang berperan penting dalam pengolahan teks adalah **Natural Language Processing (NLP)**. Dalam konteks ini, **summarization** atau teknik peringkasan otomatis merupakan salah satu pendekatan NLP yang dapat membantu meringkas tafsir tanpa kehilangan esensi dari teks aslinya. Summarization menawarkan kemudahan bagi pengguna untuk memahami inti dari tafsir yang panjang dalam waktu yang lebih singkat.

Pada umumnya, terdapat dua metode utama dalam summarization, yaitu **extractive** dan **abstractive summarization**. **Extractive summarization** mengekstraksi kalimat-kalimat penting dari teks asli, sedangkan **abstractive summarization** mencoba meringkas informasi dengan menyusun kalimat baru yang lebih singkat dan koheren. Abstractive summarization memiliki keunggulan dalam menciptakan ringkasan yang lebih alami dan tidak sekadar mengutip langsung teks asli, namun memiliki tantangan lebih besar dalam memahami konteks yang kompleks (Alami Merrouni et al., 2023). Dalam konteks teks bilingual seperti tafsir Al-Qur'an dalam bahasa Arab dan Indonesia, abstractive summarization memberikan potensi yang lebih besar untuk menghasilkan ringkasan yang informatif dan tidak bergantung pada struktur kalimat asli, sehingga cocok untuk menangani kompleksitas dan perbedaan bahasa tersebut.

Salah satu model yang dapat mendukung penerapan **abstractive summarization** adalah **BERT** (Bidirectional Encoder Representations from Transformers). BERT telah terbukti sangat efektif dalam berbagai tugas NLP, termasuk **abstractive summarization**, karena kemampuannya untuk memahami konteks dua arah secara mendalam dalam teks. Ini memungkinkan BERT untuk menghasilkan ringkasan yang lebih akurat dan koheren, terutama untuk teks panjang dan kompleks. Namun, penggunaan model BERT dalam konteks tafsir Al-Qur'an yang bilingual (Arab-Indonesia) masih belum banyak dieksplorasi. Padahal, potensinya sangat besar untuk meningkatkan aksesibilitas dan pemahaman terhadap tafsir dengan menghasilkan ringkasan otomatis yang lebih informatif dan mudah dipahami.

Berdasarkan hal tersebut, penelitian ini bertujuan untuk mengoptimalkan penggunaan metode **abstractive summarization** berbasis model transformer **BERT** dalam meringkas tafsir Al-Qur'an dalam dua bahasa, yaitu Arab dan Indonesia. Dengan memanfaatkan teknologi ini, diharapkan hasil ringkasan dapat tetap mempertahankan makna dan esensi tafsir yang kompleks, namun disajikan dalam bentuk yang lebih ringkas dan mudah dipahami oleh pengguna. Penelitian ini juga diharapkan dapat mengisi kesenjangan penelitian sebelumnya yang belum sepenuhnya memanfaatkan potensi summarization berbasis transformer untuk teks religius bilingual, khususnya dalam konteks tafsir Al-Qur'an.

1.2 Batasan Masalah

Penelitian ini memiliki beberapa batasan untuk memastikan fokus yang jelas dan pencapaian tujuan yang diinginkan. Pertama, penelitian ini difokuskan hanya pada teks tafsir Al-Qur'an yang berbahasa Arab serta terjemahannya dalam bahasa Indonesia, sehingga tidak mencakup teks keagamaan lain atau teks tafsir dalam bahasa selain Arab dan Indonesia. Kedua, penelitian ini menggunakan metode **abstractive summarization** berbasis transformer **BERT**, tanpa menerapkan metode lain di luar pendekatan ini. Model yang digunakan terbatas pada BERT (Bidirectional Encoder Representations from Transformers) yang disesuaikan untuk menangani teks bilingual (Arab-Indonesia), dan tidak mengeksplorasi model lain seperti GPT, T5, atau model transformer lainnya.

Selain itu, evaluasi kinerja dilakukan menggunakan metrik standar seperti ROUGE, tanpa menggunakan metrik evaluasi lain yang lebih kompleks atau spesifik terhadap domain tertentu. Dataset yang digunakan juga dibatasi pada kumpulan tafsir yang telah tersedia secara publik dan terjemahannya yang resmi, sehingga tidak dilakukan pengumpulan atau pembuatan dataset baru dari sumber yang belum terverifikasi. Akhirnya, penelitian ini hanya mempertimbangkan konteks teks pada level kalimat dan paragraf dalam tafsir, tanpa memasukkan konteks historis atau sosial dari tafsir Al-Qur'an secara keseluruhan.

1.3 Rumusan Masalah

Berdasarkan latar belakang dan batasan yang telah dijelaskan, rumusan masalah dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana menerapkan metode **hybrid summarization** berbasis transformer BERT untuk menghasilkan ringkasan otomatis dari teks bilingual (Arab-Indonesia) pada tafsir Al-Qur'an?
2. Sejauh mana pendekatan hybrid yang menggabungkan metode extractive dan abstractive summarization mampu menangkap esensi dari tafsir Al-Qur'an dalam dua bahasa?
3. Bagaimana performa model hybrid summarization dalam meringkas tafsir Al-Qur'an bila dievaluasi menggunakan metrik seperti ROUGE?

1.4 Tujuan Penelitian

Tujuan penelitian ini dapat dirinci sebagai berikut:

1. Menerapkan metode *abstraction summarization* berbasis transformer BERT untuk menghasilkan ringkasan otomatis dari teks bilingual (Arab-Indonesia) pada tafsir Al-Qur'an.
2. Sejauh mana pendekatan **abstractive summarization** mampu menangkap esensi dari tafsir Al-Qur'an dalam dua bahasa?.
3. Bagaimana performa model **abstractive summarization** dalam meringkas tafsir Al-Qur'an bila dievaluasi menggunakan metrik seperti ROUGE?

1.5 Manfaat Penelitian

1. Kontribusi Teoritis:

- Mengembangkan dan memperkaya metode **abstractive summarization** berbasis transformer BERT untuk teks bilingual (Arab-Indonesia), khususnya dalam konteks tafsir Al-Qur'an.
- Menyediakan wawasan baru dalam literatur pemrosesan bahasa alami (NLP) dan summarization dengan penerapan teknologi canggih pada teks keagamaan.

2. Manfaat Praktis:

- Mempermudah proses meringkas teks tafsir Al-Qur'an, membuat informasi lebih aksesibel dan mudah dipahami oleh pembaca bilingual.
- Menyediakan model summarization yang dapat digunakan oleh lembaga penelitian, pengembang aplikasi keagamaan, dan institusi pendidikan untuk menyederhanakan informasi dari tafsir dan meningkatkan kualitas bahan ajar serta sumber daya belajar.

3. Aplikasi Potensial:

- Model yang dikembangkan dapat diterapkan dalam aplikasi edukasi dan penelitian untuk memfasilitasi pemahaman yang lebih baik terhadap tafsir Al-Qur'an.
- Meningkatkan efisiensi dalam pengembangan sumber daya belajar dan materi ajar yang berkaitan dengan teks keagamaan.

1.6 Kebaruan Penelitian

Kebaruan penelitian ini terletak pada penerapan **abstractive summarization** berbasis transformer **BERT** dalam meringkas teks bilingual (Arab-Indonesia) pada tafsir Al-Qur'an. Dari segi konsep, penelitian ini memperkenalkan penggunaan pendekatan **abstractive** untuk meringkas teks keagamaan, khususnya tafsir Al-Qur'an, yang belum banyak dieksplorasi. Pendekatan ini menawarkan cara baru dalam menangani teks religius yang kompleks dan bilingual, yang bertujuan menghasilkan ringkasan yang lebih alami dan koheren. Dari segi metode, penelitian ini mengadopsi transformer **BERT** yang disesuaikan untuk menangani teks dalam dua bahasa secara bersamaan. Hal ini memungkinkan BERT untuk memahami

konteks bahasa Arab dan Indonesia secara efektif, sehingga dapat menghasilkan ringkasan yang lebih akurat dan informatif. Inovasi ini memperluas cakupan aplikasi model transformer dalam pemrosesan teks religius bilingual, memberikan metode baru dalam menghasilkan ringkasan otomatis yang efektif dan relevan.

BAB II

Kajian Pustaka

2.1 Tinjauan Pustaka

Pada penelitian mengenai summarization berbasis **abstractive summarization** menggunakan model **BERT** untuk teks bilingual (Arab-Indonesia) dalam konteks tafsir Al-Qur'an, sejumlah penelitian sebelumnya dapat menjadi landasan penting. Summarization terbagi dalam dua metode utama: **extractive** dan **abstractive**. Dalam extractive summarization, kalimat penting diambil dari teks asli tanpa perubahan struktur, sementara abstractive summarization menghasilkan kalimat baru berdasarkan pemahaman semantik terhadap teks. Penelitian oleh (Nallapati et al., n.d.) menunjukkan efektivitas metode extractive dalam menghasilkan ringkasan cepat, tetapi dengan keterbatasan dalam fleksibilitas dan akurasi semantik. Sebaliknya, penelitian (Liu & Lapata, n.d.) tentang abstractive summarization berbasis BERT memperlihatkan kemampuan model ini dalam menghasilkan ringkasan yang lebih mendekati interpretasi manusia.

Penelitian oleh (Alselwi & Taşcı, 2024) mengembangkan teknik summarization berbasis graf untuk teks Arab menggunakan **word embedding** dan algoritma **PageRank**, yang dikenal sebagai **GEATS (Graph-based Extractive Arabic Text Summarization)**. Teknik ini memanfaatkan kekuatan pendekatan berbasis graf dalam menangani ciri khas bahasa Arab, seperti hubungan morfologis yang kompleks. Hasil penelitian mereka menunjukkan bahwa pendekatan GEATS mampu menghasilkan ringkasan yang lebih berkualitas dibandingkan metode mutakhir lainnya, dengan peningkatan lebih dari 7,5% dalam nilai **F-measure**. Namun, penelitian ini berfokus pada summarization ekstraktif untuk teks monolingual Arab, tanpa memasukkan konteks tafsir Al-Qur'an atau terjemahan bilingual. Hal ini menunjukkan masih adanya peluang untuk mengembangkan model yang lebih spesifik untuk teks keagamaan dan bilingual, seperti yang ditargetkan dalam penelitian ini.

2.2 Landasn Teoritik

2.3 Kerangka Berpikir

2.4 Hipotesis Teoritis

DAFTAR PUSTAKA

- Aji Fitra Jaya Institut Perguruan Tinggi Ilmu Al Qur, S. (n.d.). *AL-485 ¶\$IIDAN HADIS SEBAGAI SUMBER HUKUM ISLAM*.
- Alami Merrouni, Z., Frikh, B., & Ouhbi, B. (2023). EXABSUM: a new text summarization approach for generating extractive and abstractive summaries. *Journal of Big Data*, 10(1). <https://doi.org/10.1186/s40537-023-00836-y>
- Alselwi, G., & Taşcı, T. (2024). Extractive Arabic Text Summarization Using PageRank and Word Embedding. *Arabian Journal for Science and Engineering*. <https://doi.org/10.1007/s13369-024-08890-1>
- Liu, Y., & Lapata, M. (n.d.). *Text Summarization with Pretrained Encoders*. <https://github.com/>
- Nallapati, R., Zhou, B., dos Santos, C., & Xiang, B. (n.d.). *Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond*.