



**Pengembangan Model BERT dengan Metode Hybrid
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 Hybrid 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 utama ajaran agama. 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, ada dua metode utama dalam summarization, yaitu **extractive** dan **abstractive summarization**. Extractive summarization mengekstraksi kalimat-kalimat penting dari teks asli, sementara abstractive summarization mencoba merangkum informasi dengan menyusun kalimat baru yang lebih singkat. Meski keduanya memiliki keunggulan masing-masing, keduanya juga memiliki keterbatasan, terutama dalam konteks teks bilingual yang kompleks seperti tafsir Al-Qur'an dalam bahasa Arab dan Indonesia. Oleh karena itu, pendekatan **hybrid summarization**, yang menggabungkan kedua metode tersebut, mulai menarik perhatian dalam upaya untuk menciptakan ringkasan yang lebih akurat dan koheren.

Salah satu model yang dapat mendukung penerapan hybrid summarization adalah **BERT (Bidirectional Encoder Representations from Transformers)**. BERT telah terbukti efektif dalam menangani berbagai tugas NLP, termasuk summarization, karena kemampuan model ini untuk memahami konteks dua arah dalam teks, terutama dalam teks panjang dan kompleks. Namun, penggunaan model BERT dalam konteks tafsir Al-Qur'an yang bilingual (Arab-Indonesia) masih belum banyak dijelajahi, meskipun potensinya cukup besar dalam meningkatkan aksesibilitas dan pemahaman terhadap tafsir.

Berdasarkan hal tersebut, penelitian ini bertujuan untuk mengoptimalkan penggunaan metode hybrid summarization berbasis model transformer BERT dalam meringkas tafsir Al-Qur'an dalam dua bahasa, yaitu bahasa Arab dan Indonesia. Dengan memanfaatkan teknologi ini, diharapkan hasil ringkasan 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 yang bilingual.

1.2 Batasan Masalah

Penelitian ini memiliki beberapa batasan untuk memastikan fokus yang jelas dan pencapaian tujuan yang diinginkan. Pertama, penelitian hanya difokuskan 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 **hybrid summarization** yang menggabungkan pendekatan extractive dan abstractive summarization berbasis transformer BERT, dengan tidak menerapkan metode lain di luar pendekatan hybrid. Model yang digunakan terbatas pada **BERT** (Bidirectional Encoder Representations from Transformers) yang disesuaikan untuk menangani teks bilingual (Arab-Indonesia), tanpa 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 hybrid summarization berbasis transformer BERT untuk menghasilkan ringkasan otomatis dari teks bilingual (Arab-Indonesia) pada tafsir Al-Qur'an.

2. Mengevaluasi efektivitas teknik hybrid yang menggabungkan pendekatan extractive dan abstractive dalam menangkap esensi dan makna dari tafsir Al-Qur'an.
3. Menilai performa model summarization yang diusulkan dengan menggunakan metrik evaluasi seperti ROUGE untuk memastikan kualitas ringkasan yang dihasilkan.

1.5 Manfaat Penelitian

1. Kontribusi Teoritis:

- Mengembangkan dan memperkaya metode **hybrid 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 **hybrid summarization** berbasis transformer BERT dalam meringkas teks bilingual (Arab-Indonesia) pada tafsir Al-Qur'an. Dari segi konsep, penelitian ini memperkenalkan penggunaan pendekatan hybrid dalam summarization yang menggabungkan metode **extractive** dan **abstractive** pada teks keagamaan, khususnya tafsir Al-Qur'an, yang sebelumnya belum banyak dieksplorasi. Pendekatan ini menawarkan cara baru dalam menangani teks kompleks dan bilingual, serta memberikan kontribusi penting dalam bidang pemrosesan bahasa alami (NLP). Dari segi metode, penelitian ini mengadopsi transformer BERT yang disesuaikan untuk menangani teks dalam dua bahasa secara bersamaan, sehingga mampu menghasilkan ringkasan yang lebih

informatif dan akurat. Inovasi ini mencakup penerapan teknologi canggih dalam summarization pada teks religius dan bilingual, memperluas cakupan aplikasi model transformer dan memberikan metode baru dalam menghasilkan ringkasan yang efektif dan relevan.