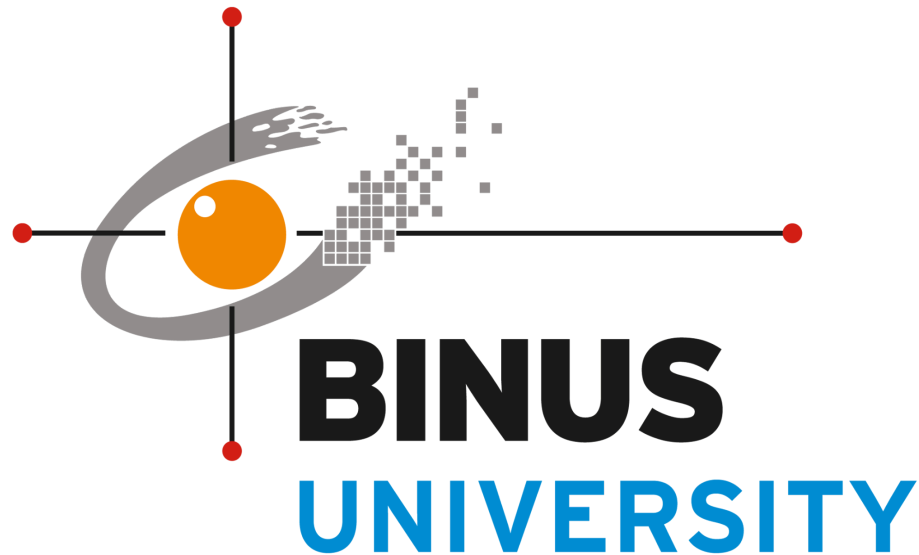


Text Summarize Model for Daily News



Group 11

Brian Alexander - 2702282351

Jason Christian Buddhihartono - 2702326593

Melvorn Michio Chie - 2702220946

Kelas : LE01

Mata Kuliah : Deep Learning

Dosen Pengampu:

Ricky Reynardo Siswanto, S.T., M.Kom.

Tahun Ajaran 2025/2026

1. INTRODUCTION

Perkembangan sains dan tekno dalam beberapa tahun terakhir mengalami peningkatan yang sangat pesat. Salah satu bidang teknologi yang berkembang signifikan adalah *Artificial Intelligence* (AI). *Large Language Model* (LLM) telah membuka peluang baru dalam pengolahan bahasa alami, khususnya dalam kemampuan membaca teks, memahami konteks dari teks yang diberikan, serta merangkum informasi tersebut secara bersamaan dalam waktu yang lebih cepat dibandingkan kemampuan manusia pada umumnya

Sistem LLM yang berfungsi sebagai *generator* AI mampu membaca lebih dari dua dokumen secara simultan, mengolah data berupa teks dari dokumen-dokumen tersebut, serta mengekstraksi informasi penting yang merepresentasikan poin-poin utama dari seluruh dokumen yang dianalisis. Selain itu, sistem ini juga menghasilkan teks ringkasan yang mampu menjelaskan isi dokumen secara lebih ringkas, jelas, dan terstruktur.

Pemanfaatan teknologi yang terus berkembang ini oleh masyarakat umum berfokus pada model AI yang memiliki kemampuan pengambilan konteks dan pembangkitan bahasa. Teknologi tersebut dapat membantu pengguna dalam memproses dokumen secara lebih efisien dan mempermudah berbagai aktivitas sehari-hari. Dampak dari perkembangan ini dirasakan oleh banyak pihak, terutama di lingkungan perkantoran yang sebagian besar aktivitasnya berkaitan dengan pengolahan dokumen, sehingga penggunaan LLM dapat memberikan manfaat yang signifikan dan meningkatkan produktivitas.

Pendekatan model ini terhadap generasi teks sangat berguna dalam meningkatkan efisiensi analisis informasi, khususnya dalam kegiatan penelitian. Pada berbagai skenario yang melibatkan data teks dalam jumlah besar, seperti artikel, laporan saintifik, maupun dokumen akademik atau journal, LLM mampu membantu proses pemahaman dan peringkasan informasi secara efektif

2. DATASET

Dataset diperoleh dari Kaggle dengan judul "*CNN-DailyMail News Text Summarization*" yang berisi 300 ribu berita CNN dalam bahasa Inggris. Dataset terbagi menjadi 3 bagian yaitu train untuk pelatihan model, validation untuk validasi model saat proses pelatihan, dan test untuk evaluasi akhir model setelah proses pelatihan selesai. Dataset sendiri dapat diakses pada link berikut:

<https://www.kaggle.com/datasets/gowrishankarp/newspaper-text-summarization-cnn-dailymail>

3. MODELLING

3.1 Model

Model yang digunakan pada project ini adalah **Flan-T5** versi base, varian dari model T5 (Text-to-Text Transfer Transformer) milik Google. Tidak seperti metode ekstraksi sederhana yang hanya mengambil kata yang penting, Flan-T5 menggunakan pendekatan abstraktif. Dimana model ini membaca teks secara menyeluruh dan akan men generate kalimat yang baru dengan makna utama yang difokuskan dan meniru bagaimana manusia meringkas sebuah teks.

Flan-T5 menggunakan arsitektur transformer untuk encoder maupun decodernya. Membedakan model ini dengan model GPT lainnya yang hanya mengandalkan decoder pada transformer saja. Encoder sendiri akan memproses teks secara berurutan dari kiri maupun kanan (bidirectional) sehingga menciptakan vektor konteks, sebuah representasi matematika yang menangkap arti dari teks tersebut, memberikan attention pada setiap kata yang berhubungan (self-attention). Kemudian decoder akan mengambil hasil vektor dari encoder dan menghasilkan teks kembali satu kata demi satu kata. Decoder juga menggunakan mekanisme Cross-Attention untuk melihat kembali teks asli sambil generate summary untuk memastikan bahwa hasilnya akan tetap benar dan sesuai dengan teks aslinya.

Untuk Flan sendiri berasal dari Finetuned Language Net, model yang lebih canggih dari T5 telah dilatih lebih lanjut untuk tugas peringkas sehingga mengurangi waktu dan kebutuhan dataset yang banyak. Di project ini, kami menggunakan versi base karena parameternya yang lebih sedikit yaitu 250M parameter sehingga tidak membebankan proses pelatihan serta keterbatasan resource yang dimiliki.

Berikut beberapa keunggulan dari flan-t5 yang menjadi pertimbangan kami dalam memilih model ini:

- **Sudah instructed fine-tuned**
Flan-t5 telah di fine-tuned untuk perintah meringkas sehingga lebih stabil dan akurat dan cocok untuk project ini karena fokus pada tugas khusus (meringkas).
- **Banyak digunakan dalam research**
Flan-t5 sendiri sudah banyak research dan pembahasan terkait model ini sehingga proses pelatihan model terarah.
- **GPU Efficiency**
Banyak varian dari flan-t5 dengan parameter yang lebih kecil sehingga beban komputasi lebih kecil namun performa model tetap stabil

3.2 Setup Training

Dataset yang digunakan pada project ini hanya sebanyak 50.000 data teks berita dan ringkasannya untuk data training dan 10.000 untuk data testing. Pengurangan pada dataset dikarenakan terbatasnya resource komputasi untuk pelatihan model untuk mengurangi waktu training model T5. Pengurangan tersebut tidak terlalu signifikan karena pada proses uji coba, kami membandingkan pelatihan data 20.000 dengan 50.000 dan hasilnya tidak terlalu berbeda.

Arsitektur model sendiri kami set input sebesar maksimal 1028 token dan output 128 token sehingga komputasi tidak berat dan menjaga kualitas teks pada input dan menjaga hasil output yang ringkas mungkin namun tetap memberikan arti pada keseluruhan teks tersebut. Kami tidak banyak mengubah arsitektur model secara general, hanya beberapa hyperparameter training seperti weight decay dan learning yang kami tentukan sebelum melakukan proses pelatihan.

4. EVALUATION

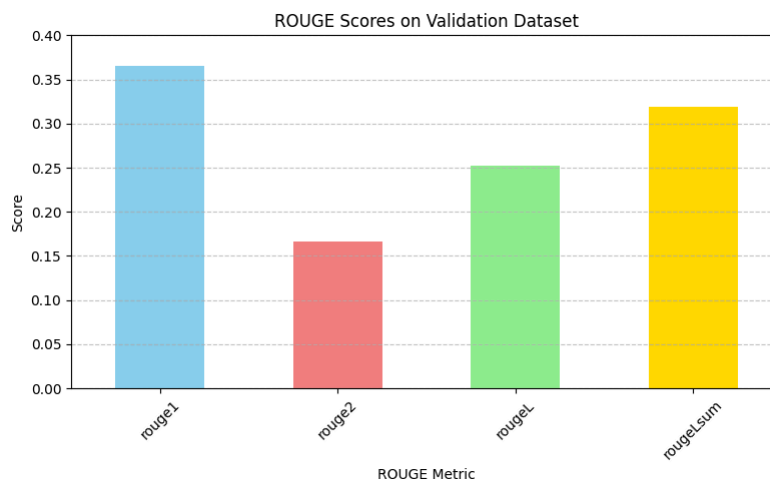
Untuk mengukur efektivitas model *flan-t5* yang telah dilatih (*fine-tuned*), kami menggunakan metrik evaluasi standar industri untuk proyek *text summarization*, yaitu **ROUGE** (*Recall-Oriented Understudy for Gisting Evaluation*). Metrik ini bekerja dengan cara membandingkan ringkasan yang dihasilkan model (*generated summary*) dengan ringkasan referensi yang dibuat oleh manusia (*reference summary*) dari dataset validasi.

Evaluasi difokuskan pada empat varian skor ROUGE untuk mendapatkan gambaran performa yang menyeluruh:

- **ROUGE-1**: Mengukur tumpang tindih (*overlap*) dari *unigram* (kata tunggal). Skor ini penting untuk menilai seberapa banyak informasi atau kata kunci penting dari teks asli yang berhasil ditangkap oleh model.
- **ROUGE-2**: Mengukur tumpang tindih dari *bigram* (pasangan dua kata berurutan). Metrik ini lebih ketat dan digunakan untuk menilai kelancaran dan ketepatan frasa yang dihasilkan.
- **ROUGE-L**: Mengukur *Longest Common Subsequence* (LCS). Metrik ini menilai struktur kalimat dengan mencari urutan kata terpanjang yang sama antara hasil model dan referensi tanpa harus berurutan secara sempurna, yang merefleksikan koherensi tingkat kalimat.
- **ROUGE-Lsum**: Mirip dengan ROUGE-L, namun dihitung pada level ringkasan secara keseluruhan (antara kalimat), bukan hanya per kalimat. Ini memberikan gambaran berkualitas ringkasan secara utuh.

Hasil Evaluasi Kuantitatif

Berdasarkan proses validasi, berikut adalah visualisasi skor ROUGE model kami:



Seperti yang terlihat pada grafik di atas, model menunjukkan performa yang cukup menjanjikan untuk ukuran model yang cukup ringan (250M parameter):

1. **ROUGE-1** mencatat skor tertinggi di angka sekitar **0.36**. Hal ini mengindikasikan bahwa model sangat kapabel dalam mengidentifikasi dan mempertahankan kata-kata kunci utama dari berita yang diringkas.
2. **ROUGE-Lsum** berada di posisi kedua dengan skor mendekati **0.32**, yang menandakan bahwa secara keseluruhan, ringkasan yang dihasilkan memiliki struktur yang menyerupai referensi manusia dan cukup informatif.
3. **ROUGE-L** memperoleh skor sekitar **0.25**, menunjukkan kemampuan model dalam menyusun kalimat yang koheren secara sintaksis/
4. **ROUGE-2** berada di angka sekitar **0.16**. meskipun rendah dibandingkan metrik-metrik lainnya, hal ini cukup wajar dalam peringkasan abstraktif, karena model dituntut untuk menyusun ulang kalimat dengan gaya bahasa model itu sendiri, sehingga kemungkinan kecocokan frasa dua kata (bigram) yang persis sama dengan referensi manusia menjadi lebih kecil.

Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa model Flan-T5 yang kami gunakan mampu menghasilkan ringkasan berita yang relevan dan dapat dimengerti, menyeimbangkan antara pengambilan informasi penting (ROUGE-1) dan keterbacaan struktur (ROUGE-Lsum).

5. DEPLOYMENT

Model yang sudah dilatih kemudian dideploy pada aplikasi streamlit yang berbasis python. Streamlit dipilih karena model yang dilatih juga menggunakan python sehingga memudahkan proses import model dan prediksi. Kemudian Streamlit juga sangat mudah digunakan dan sudah mendukung desain website sederhana sehingga pengujian model bisa dilakukan dengan cepat. Di website ini, user bisa langsung mencoba teks yang ingin mereka ringkas, adapun template teks yang kami sediakan sehingga user bisa mencoba dan melihat hasil dari model peringkasan yang digunakan.

6. REFLECTION

Melalui project ini, kami belajar banyak hal baik hard skill maupun soft skill. Kami diharuskan untuk bisa bekerja sama dengan baik dan memiliki koordinasi yang baik agar project bisa berjalan dengan lancar. Selain itu juga, kami belajar banyak hal terkait dengan project kami, khususnya model peringkasan teks ini. Kami jadi mengerti bagaimana model T5 ini meringkas sebuah teks dengan memanfaatkan encoder, decoder, dan mekanisme attention agar hasil ringkasannya bisa bagus dan sesuai dengan teks orisinal. Secara keseluruhan, kami bersyukur project ini bisa selesai dengan baik dan kami banyak belajar hal-hal saat pembuatan project deep learning ini yang dapat bermanfaat bagi kami.

REFERENCES

- <https://arxiv.org/abs/2210.11416>
- <https://arxiv.org/abs/2108.08877>