

TEXT SUMMARIZER

BRIAN ALEXANDER - 2702282351

JASON CHRISTIAN BUDHIHARTONO - 2702326593

MELVERN MICHIO CHIE - 2702220946

INTRODUCTION

- Keberadaan AI sangat membantu kita dalam menyelesaikan permasalahan sehari-hari
- Munculnya LLM memungkinkan text generation
- Berita yang panjang sangat memakan waktu untuk dibaca seluruhnya
- Butuh model peringkas teks untuk membantu





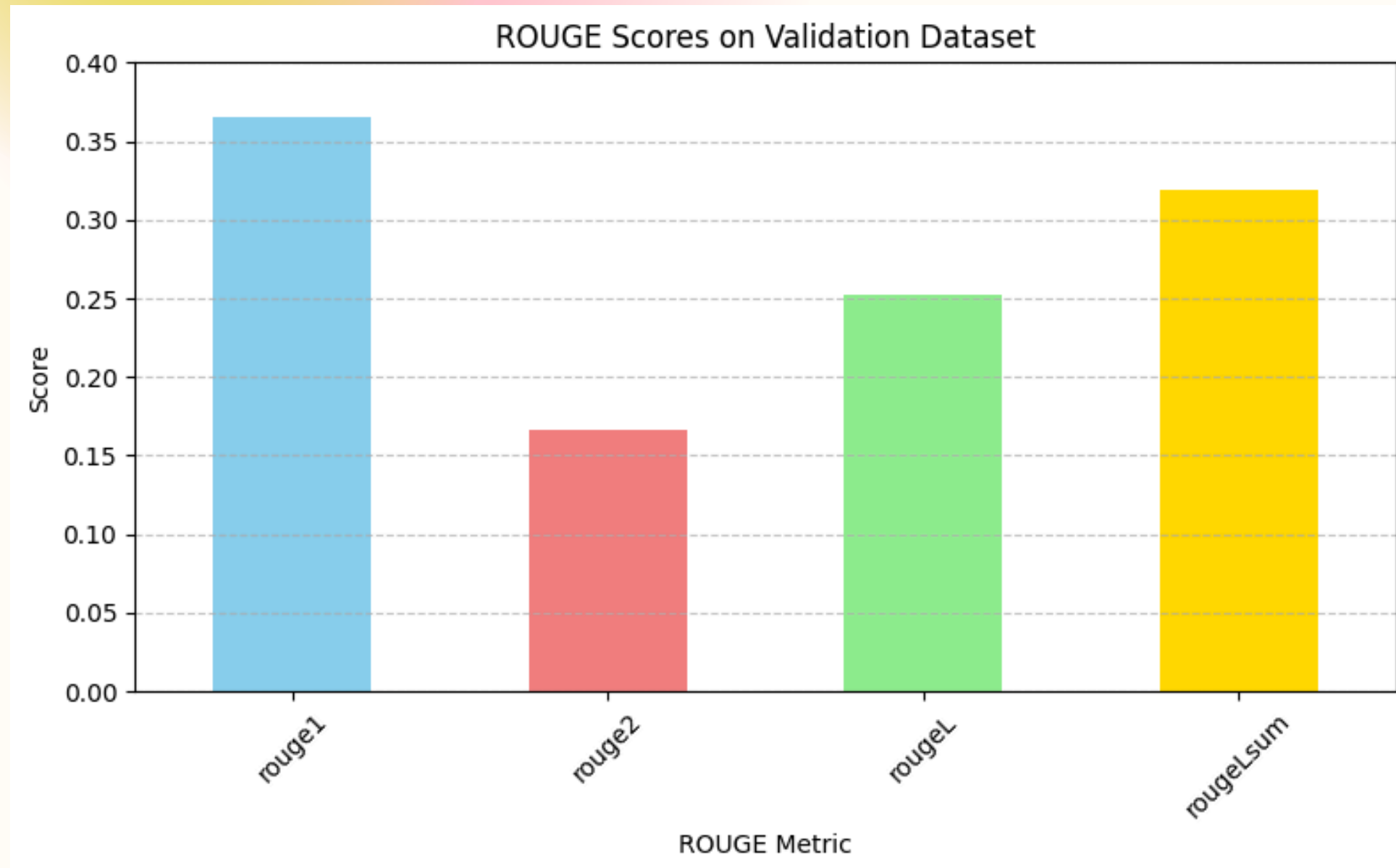
DATASET

Dataset diperoleh dari Kaggle dengan judul “CNN-DailyMail News Text Summarization” . Dataset sendiri dapat diakses pada link berikut: <https://www.kaggle.com/datasets/gowrishankarp/newspaper-text-summarization-cnn-dailymail>

MODELLING

- Model yang digunakan: **Flan-T5 base (Google)**
- **Flan** → sudah finetune lebih jauh untuk task meringkas
- Menggunakan **Encoder** dan **Decoder**
- **Encoder** → memproses teks secara berurutan (Bidirectional) dan menciptakan representasi berupa vektor
- **Decoder** → Mengambil hasil vektor per kata sambil generate summary dan memastikan hasil sesuai dengan teks original

EVALUATION



Hasil evaluasi ROUGE (Recall-Oriented Understudy for Gisting Evaluation) 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 kalimat (ROUGE-Lsum).

DEMO

THANK
YOU