

BAB I

PENDAHULUAN

1.1. Latar Belakang

Penggunaan video sebagai media penyampaian informasi publik terus mengalami peningkatan, terutama melalui platform daring seperti YouTube [1]. Akses yang mudah dan luas membuat video menjadi pilihan utama dalam menyampaikan informasi, baik untuk kebutuhan hiburan, pendidikan, maupun penyampaian kebijakan publik. Namun, di balik kemudahan tersebut, video memiliki keterbatasan terutama terkait dengan kecepatan konsumsi informasi. Durasi yang panjang serta struktur konten yang tidak selalu sistematis dapat menjadi hambatan bagi audiens untuk memperoleh inti informasi dengan cepat [2]. Hal ini menjadi persoalan penting, terutama dalam konteks video yang memuat pembelajaran, pidato kebijakan, atau diskusi publik yang panjang, di mana masyarakat membutuhkan pemahaman cepat terhadap informasi yang disampaikan.

Kondisi tersebut meningkatkan potensi terjadinya kesalahpahaman, salah interpretasi, hingga hilangnya informasi penting, terutama bagi audiens dengan keterbatasan waktu atau akses. Dengan demikian, sebuah metode untuk mengekstraksi inti informasi dari video menjadi semakin relevan. Salah satu pendekatan yang umum digunakan adalah mentranskripsikan audio dalam video ke dalam bentuk teks melalui teknologi speech-to-text. Kemajuan teknologi seperti Whisper telah memungkinkan konversi audio ke teks dengan akurasi yang cukup tinggi [3], [4]. Akan tetapi, hasil transkripsi tersebut biasanya masih berupa teks repetit, repetitive, dan tidak terorganisasi dengan baik sehingga belum bisa langsung dimanfaatkan sebagai ringkasan yang padat dan informatif. Selain itu, transkripsi yang berasal dari komunikasi lisan cenderung tidak memiliki alur naratif yang jelas dan rapi, sehingga ringkasan harus mampu menangkap keragaman linguistik tersebut agar tetap relevan dan mudah dipahami.

Dalam bidang pemrosesan bahasa alami (Natural Language Processing), teknik *text summarization* merupakan pendekatan utama untuk menyederhanakan konten transkripsi video. Secara umum, summarization terbagi menjadi dua pendekatan: ekstraktif dan abstraktif. Pendekatan ekstraktif memilih kalimat-kalimat penting dari teks asli, sedangkan pendekatan abstraktif menghasilkan kalimat baru yang lebih ringkas dan koheren dengan memanfaatkan pemahaman konteks secara mendalam.

BERT adalah salah satu metode yang telah banyak digunakan pada proses extractive summarization karena kemampuannya dalam membangun representasi kalimat berbasis konteks secara mendalam [5], [6]. BERT mampu menilai kesemantisan antar kalimat, sehingga memudahkan dalam pemilihan kalimat yang relevan untuk dijadikan ringkasan. Namun, pemilihan kalimat penting dari teks transkripsi yang panjang tetap menghadapi tantangan lain, seperti adanya redundansi informasi, pengulangan topik, serta perubahan topik yang cepat. Untuk mengatasi hal tersebut, penelitian ini menerapkan algoritma *k-means clustering* guna mengelompokkan kalimat berdasarkan kesamaan semantik sehingga dapat dipilih kalimat representatif dari setiap topik [7]. Meskipun pendekatan ekstraktif efektif untuk mereduksi panjang teks, hasilnya kadang terasa kaku karena hanya mengutip kalimat apa adanya. Oleh sebab itu, penelitian ini menambahkan tahap *abstractive summarization* menggunakan model BART.

BART, dengan arsitektur encoder–decoder yang menggabungkan kemampuan *bidirectional encoding* dan *autoregressive decoding*, mampu memahami konteks global dari teks masukan sekaligus menghasilkan kalimat baru yang lebih terstruktur dan alami [8]. Pada tahap encoder, BART memproses kalimat-kalimat hasil ekstraksi secara menyeluruh sehingga mampu menangkap hubungan antar-kalimat, dependensi semantik, serta alur makna dari teks yang sebelumnya tidak teratur. Sementara itu, pada tahap decoder, BART menyusun ulang informasi tersebut menjadi rangkasan yang lebih ringkas, koheren, dan mudah dipahami, bukan sekadar menyalin atau memotong kalimat asli [9].

Berdasarkan uraian tersebut, penelitian ini mengusulkan metode hibrida yang menggabungkan *extractive summarization* berbasis BERT dan *k-means clustering*, kemudian dilanjutkan tahap *abstractive summarization* menggunakan BART untuk menghasilkan ringkasan transkripsi video yang lebih koheren dan natural.. Pendekatan ini diterapkan pada kumpulan data berupa transkripsi dari beberapa video publik di YouTube yang mencakup berbagai jenis komunikasi nyata seperti monolog, wawancara, dan podcast. Evaluasi kualitas ringkasan dilakukan dengan membandingkan hasil ringkasan otomatis terhadap ringkasan manual menggunakan metrik ROUGE. Melalui penelitian ini diharapkan dapat diperoleh metode ringkasan transkripsi video yang efektif, informatif, serta mampu menjawab tantangan penyajian informasi dalam format video berdurasi panjang.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, masalah penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana kinerja pendekatan hibrida yang menggabungkan ekstraksi berbasis BERT dan pengelompokan *K-Means*, serta abstraksi menggunakan BART, dalam merangkum transkripsi video dengan berbagai jenis konten seperti monolog, wawancara, dan podcast?
2. Bagaimana variasi struktur video dan keragaman topik memengaruhi kualitas ringkasan otomatis, serta apa keterbatasan yang muncul dalam penerapan metode ini pada percakapan kompleks dan dataset bertopik tunggal?

1.3. Tujuan Penelitian

Sejalan dengan latar belakang dan rumusan masalah yang telah dipaparkan sebelumnya, penelitian ini memiliki beberapa tujuan yang hendak dicapai sebagai landasan dalam pengembangan metode peringkasan otomatis. Tujuan dari penelitian ini adalah:

1. Menganalisis kinerja pendekatan hibrida yang mengombinasikan metode ekstraktif berbasis BERT dan pengelompokan *K-Means* dengan metode

abstraktif menggunakan model BART dalam menghasilkan ringkasan transkripsi video pada berbagai jenis konten seperti monolog, wawancara, dan podcast.

2. Mengevaluasi pengaruh variasi struktur video dan keragaman topik terhadap kualitas ringkasan otomatis, serta mengidentifikasi keterbatasan metode dalam menangani percakapan kompleks dan dataset dengan topik yang homogen.

1.4. Batasan Penelitian

Agar penelitian ini lebih terarah, fokus, serta tidak melebar pada aspek yang berada di luar lingkup kajian, maka diperlukan penetapan batasan yang jelas. Batasan penelitian ini adalah sebagai berikut:

1. Penelitian hanya berfokus pada metode hibrida yang menggabungkan ekstraktif (BERT + k-means) dan abstraktif (BART).
2. Dataset yang digunakan berupa transkripsi dari dua puluh video publik YouTube yang mencakup jenis komunikasi berbeda (monolog, wawancara, dan podcast).
3. Model representasi kalimat yang digunakan adalah BERT, sedangkan proses abstraktif menggunakan BART.
4. Evaluasi ringkasan dilakukan menggunakan metrik ROUGE tanpa melibatkan penilaian kualitatif dari ahli bahasa.