

BAB I

PENDAHULUAN

1.1. Latar Belakang

Perkembangan platform digital berbasis video, seperti YouTube, telah menjadikan video sebagai salah satu media utama dalam penyampaian informasi, edukasi, dan diskusi public[1],[2]. Berbagai topik penting, mulai dari kesehatan, kebijakan publik, hingga edukasi ilmiah, kini banyak disampaikan dalam bentuk video berdurasi panjang seperti monolog, wawancara, dan podcast [3]. Meskipun kaya informasi, format video memiliki keterbatasan dari sisi efisiensi konsumsi informasi. Audiens sering kali membutuhkan waktu yang cukup lama untuk memahami inti pembahasan, terutama ketika struktur penyampaian bersifat lisan, tidak sistematis, dan mengandung banyak pengulangan [4].

Untuk mengatasi keterbatasan tersebut, teknologi Automatic Speech Recognition (ASR) memungkinkan konten audio dalam video diubah menjadi bentuk teks melalui proses transkripsi [5]. Namun, hasil transkripsi video umumnya masih berupa teks mentah yang panjang, repetitif, dan tidak terstruktur dengan baik [6]. Karakteristik bahasa lisan seperti pengulangan, jeda, serta perpindahan topik yang cepat menyebabkan teks transkripsi sulit dipahami secara langsung, sehingga dibutuhkan proses lanjutan untuk merangkum informasi penting yang terkandung di dalamnya [7].

Dalam bidang Natural Language Processing (NLP), peringkasan teks (text summarization) menjadi pendekatan yang relevan untuk menyederhanakan transkripsi video [8]. Secara umum, metode peringkasan terbagi menjadi dua pendekatan utama, yaitu ekstraktif dan abstraktif. Metode ekstraktif memilih kalimat-kalimat penting langsung dari teks asli, sedangkan metode abstraktif menghasilkan kalimat baru yang lebih ringkas dengan mempertimbangkan konteks keseluruhan [9]. Meskipun metode ekstraktif relatif stabil dalam menjaga informasi penting, hasil ringkasannya sering kali

kurang mengalir dan masih terasa kaku. Sebaliknya, metode abstraktif mampu menghasilkan ringkasan yang lebih natural, namun berisiko menghilangkan informasi penting apabila tidak didukung oleh seleksi konten yang baik .

Untuk menjembatani keterbatasan tersebut, pendekatan hibrida yang mengombinasikan metode ekstraktif dan abstraktif banyak dikembangkan dalam penelitian peringkasan teks[10] . Pada tahap ekstraktif, representasi semantik kalimat menjadi faktor penting dalam menentukan kalimat yang relevan [11]. Model BERT (Bidirectional Encoder Representations from Transformers) telah terbukti efektif dalam membangun representasi kalimat berbasis konteks secara mendalam[10] . Namun, pada transkripsi video yang panjang dan kompleks, masih terdapat tantangan berupa redundansi informasi dan kemunculan topik yang berulang [10].

Oleh karena itu, proses ekstraktif dalam penelitian ini dilengkapi dengan teknik klusterisasi untuk mengelompokkan kalimat berdasarkan kesamaan semantik. Berbeda dengan pendekatan sebelumnya yang menggunakan K-Means, penelitian ini menerapkan klusterisasi Hierarchical sebagai pembeda utama . Klusterisasi hierarchical memiliki keunggulan dalam menangkap struktur hubungan antar kalimat secara bertingkat tanpa memerlukan penentuan jumlah kluster di awal[12]. Pendekatan ini dinilai lebih fleksibel dalam merepresentasikan variasi topik dan subtopik yang muncul secara alami dalam transkripsi video .

Hasil dari tahap ekstraktif berbasis BERT dan klusterisasi hierarchical kemudian diproses menggunakan model BART (Bidirectional and Auto-Regressive Transformers) pada tahap abstraktif . BART dengan arsitektur encoder–decoder mampu menyusun kembali informasi penting menjadi ringkasan yang lebih koheren, ringkas, dan mudah dipahami. Dengan menggabungkan kekuatan seleksi konten dari tahap ekstraktif dan kemampuan generatif dari tahap abstraktif, pendekatan hibrida ini diharapkan mampu menghasilkan ringkasan transkripsi video yang informatif sekaligus natural [13].

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana kinerja pendekatan hibrida yang menggabungkan peringkasan ekstraktif berbasis BERT dan klasterisasi hierarchial dengan peringkasan abstraktif menggunakan BART dalam merangkum transkripsi video?
2. Bagaimana pengaruh penggunaan klasterisasi hierarchial terhadap kualitas ringkasan transkripsi video pada berbagai jenis konten seperti monolog, wawancara, dan podcast?

1.3. Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Menganalisis kinerja pendekatan hibrida yang mengombinasikan metode ekstraktif berbasis BERT dan klasterisasi hierarchial dengan metode abstraktif menggunakan BART dalam menghasilkan ringkasan transkripsi video.
2. Mengevaluasi kualitas ringkasan yang dihasilkan pada berbagai jenis format video, seperti monolog, wawancara, dan podcast, berdasarkan hasil eksperimen dan metrik evaluasi yang digunakan.

1.4. Batasan Penelitian

Agar penelitian ini lebih terarah dan sesuai dengan ruang lingkup yang ditetapkan, maka batasan penelitian yang diterapkan adalah sebagai berikut:

1. Penelitian ini berfokus pada pendekatan hibrida yang mengombinasikan peringkasan ekstraktif berbasis BERT dengan klasterisasi *hierarchial*, serta peringkasan abstraktif menggunakan model BART.
2. Dataset yang digunakan berupa transkripsi dari dua puluh video publik YouTube yang mencakup berbagai jenis komunikasi, yaitu monolog, wawancara, dan podcast.

3. Representasi semantik kalimat pada tahap ekstraktif menggunakan model BERT, sedangkan proses peringkasan abstraktif dilakukan menggunakan model BART tanpa pelatihan ulang.
4. Evaluasi kualitas ringkasan dilakukan menggunakan metrik ROUGE tanpa melibatkan penilaian kualitatif dari ahli bahasa..

