BAB I PENDAHULUAN

1.1. Latar Belakang Penelitian

Media pembelajaran memiliki peran yang signifikan dalam proses belajar-mengajar. Saat ini, revolusi industri dan revolusi pendidikan berlangsung secara bersamaan, membawa dampak besar bagi kehidupan manusia. Perubahan dalam dunia industri turut memengaruhi perkembangan pendidikan, menciptakan tantangan baru bagi para pengajar untuk memanfaatkan teknologi secara optimal. Salah satu cara pemanfaatan teknologi dalam pendidikan adalah melalui penggunaan media pembelajaran. Dalam satu dekade terakhir, pembelajaran berbasis video telah menjadi tren dalam e-learning. Namun, untuk meningkatkan efektivitas pembelajaran, video harus menyajikan informasi utama dengan menghilangkan informasi tambahan atau yang tidak relevan.[1].

Dikutip dari survei Saiful Mujani Research and Consulting (SNRC) yang dilakukan pada rentang waktu 5-8 agustus 2020 terkait Pendidikan online di masa pendemi Covid-19, sebanyak 92% peserta didik mengalami banyak masalah dalam mengikuti pembelajaran yang dilakukan secara daring selama pandemi corona melanda. Persentase paling tinggi yaitu terkait dengan kurangnya bimbingan dari guru sebesar 38%. Salah satunya dampak dari hal itu tentang pemahaman materi yang kurang, dan rasa bosan yang timbul dari peserta didik serta tidak mengikuti pembelajaran saat pembelajaran dimulai.

Kecerdasan buatan (AI) dan pembelajaran mesin menghadirkan solusi inovatif untuk mengatasi tantangan ini. Teknik video summarization terbukti efektif dalam mengekstrak informasi penting dari video berdurasi panjang dalam waktu yang lebih singkat. Arsitektur transformer menawarkan potensi besar dengan memanfaatkan teknologi speech-to-text dan text summarization untuk menghasilkan ringkasan video yang berkualitas, sehingga mampu menyajikan informasi inti dari suatu video pembelajaran dan video berita[2].

Transformer merupakan arsitektur model yang tidak bergantung pada mekanisme pengulangan, melainkan sepenuhnya memanfaatkan mekanisme perhatian untuk memahami hubungan antara input dan output. Model ini memungkinkan proses paralelisasi yang lebih luas dan telah menunjukkan performa unggul dalam tugas penerjemahan berbasis urutan [3].

Model Whisper adalah model Automatic Speech Recognition (ASR) yang mampu mengenali sinyal suara dan mengkonversinya menjadi teks dengan tingkat akurasi tinggi, bahkan dalam kondisi data pelatihan berlabel yang terbatas. Model ini dirancang untuk memahami berbagai bahasa dan aksen dengan lebih baik melalui pendekatan pembelajaran multitugas. Whisper menggunakan jaringan syaraf berbasis Transformer yang telah dilatih secara ekstensif pada berbagai dataset audio untuk meningkatkan pemahaman konteks dalam transkripsi suara. Dalam prosesnya, model ini tidak hanya mengubah audio menjadi teks, tetapi juga mampu menangani berbagai tantangan seperti kebisingan latar belakang dan variasi intonasi. Melalui tahap pra-pelatihan, Whisper mempelajari pola ucapan dari data tanpa label, kemudian disempurnakan dengan data berlabel untuk meningkatkan akurasi. Pendekatan ini memungkinkan model beradaptasi dengan berbagai skenario pengenalan suara, termasuk dalam transkripsi video pembelajaran dan video berita secara lebih efektif [4].

Sementara itu, Model Long T5 berhasil mencapai performa tinggi berkat strategi pelatihan yang efektif. Evaluasi kinerjanya dilakukan menggunakan matrik seperti skor Rouge dan Bert untuk membandingkan hasil yang diperoleh. Model Long T5 yang digunakan mampu menghasilkan ringkasan teks dengan lebih akurat berkat representasi bahasa yang lebih baik dalam jaringan syaraf.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengeksplorasi dan menerapkan Arsitektur Transformer dalam peringkasan video guna menghasilkan ringkasan video pembelajaran dan video berita yang lebih efektif dan efisien. Dalam penelitian ini, digunakan model Whisper untuk fitur speech-to-text dan model Long T5 untuk fitur text summarization. Diharapkan bahwa penelitian ini dapat memberikan kontribusi yang berarti bagi perkembangan teknologi pendidikan serta menawarkan solusi praktis terhadap berbagai tantangan dalam pembelajaran berbasis video.

1.2. Perumusan Masalah Penelitian

Dari masalah yang ada di latar belakang di atas, dapat diambil rumusan masalah sebagai berikut.

1. Bagaimana cara menerapkan model Whisper dengan melakukan fine-tuning pada dataset bahasa Indonesia untuk mengonversi sinyal suara dari video menjadi teks dalam bahasa Indonesia?
2. Bagaimana cara menggunakan model Long T5 secara efektif dan efisien untuk merangkum teks dari hasil transkripsi audio?
3. Apakah performa kombinasi model Whisper dan Long T5 lebih baik dari model T5 dalam menghasilkan ringkasan dari video pembelajaran dan video berita secara efektif?

1.3. Batasan Masalah

Untuk memastikan pembahasan dalam penelitian ini lebih terfokus dan efektif, penulis menetapkan batasan topik sebagai berikut.

1. Penelitian ini akan terbatas pada penggunaan model Whisper untuk fitur speech-to-text dan model Long T5 untuk fitur text summarization. Algoritma atau model lain di luar kedua model ini tidak akan digunakan maupun dieksplorasi dalam penelitian ini.

2. Penelitian ini berfokus terhadap video pembelajaran dan video berita dan video berita dengan durasi lebih dari 15 menit.
3. Penelitian ini terbatas pada penggunaan bahasa Indonesia. Model Whisper akan dioptimalkan melalui fine-tuning dengan dataset bahasa Indonesia agar dapat mengonversi sinyal suara dari video menjadi teks dalam bahasa Indonesia.
4. Hasil dari penelitian ini berupa teks ringkasan dari video pembelajaran dan video berita.

1.4. Tujuan Penelitian

Tujuan yang ingin dicapai dari penelitian adalah :

1. Mengembangkan dan menyesuaikan model Whisper dengan dataset bahasa Indonesia agar mampu mengonversi sinyal suara dari video menjadi teks dalam bahasa Indonesia secara akurat.
2. Menerapkan model Long T5 dalam proses peringkasan teks, dengan fokus pada pembuatan ringkasan yang efektif dan efisien dari hasil transkripsi audio yang diperoleh menggunakan model Whisper.
3. Mengevaluasi kinerja kombinasi model Whisper dan Long T5 dalam menghasilkan ringkasan video pembelajaran dan video berita, serta menilai sejauh mana kombinasi ini dapat meningkatkan kualitas dan efisiensi pembelajaran berbasis video.

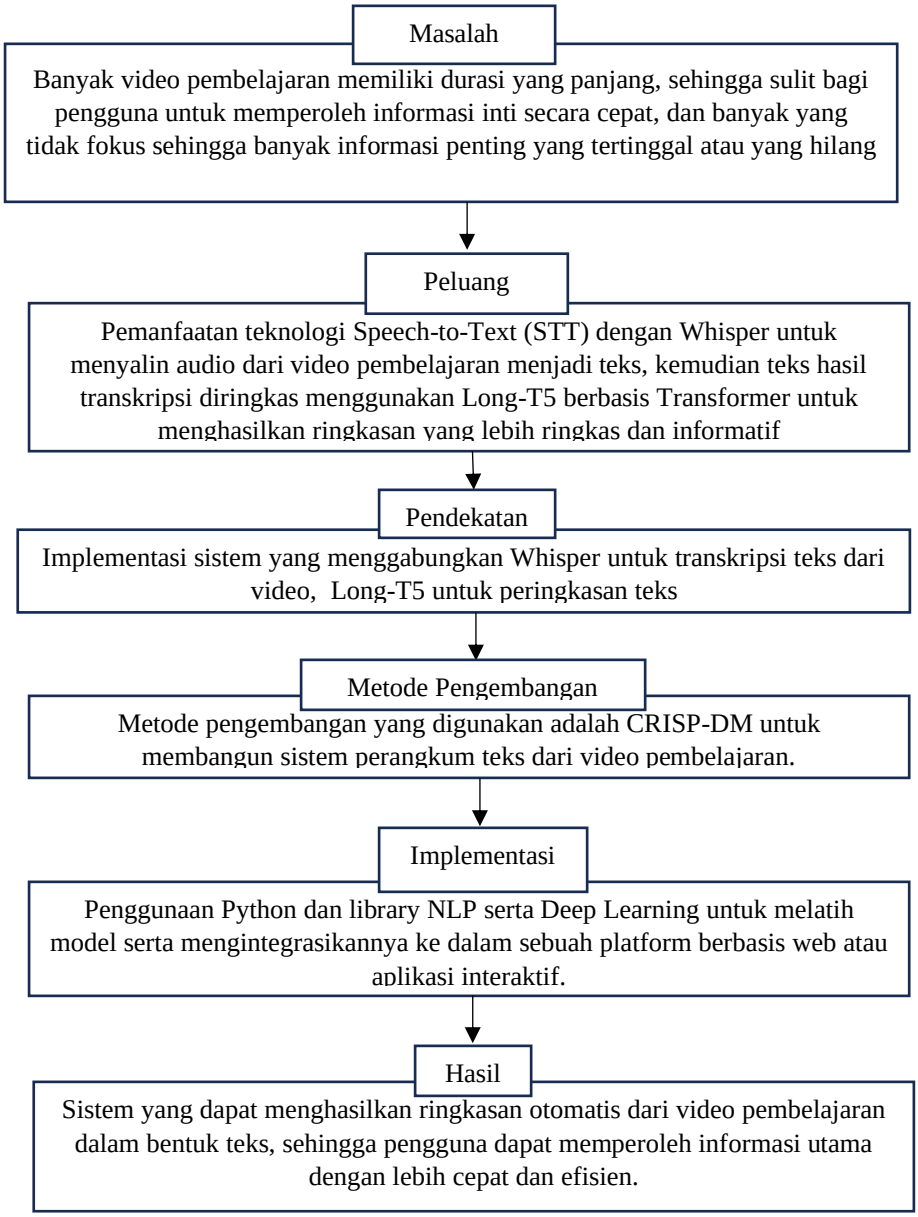
1.5. Manfaat Penelitian

Manfaat yang diharapkan tercapai dari Penelitian ini adalah :

1. Memperluas wawasan dan pemahaman mengenai penerapan arsitektur Transformer melalui penggunaan model Whisper dan Long T5 dalam bidang pengenalan suara dan peringkasan teks.
2. Meningkatkan efisiensi dan efektivitas dalam memahami konten video pembelajaran dan video berita dengan menyediakan ringkasan teks yang ringkas dan mudah dipahami.
3. Mengembangkan metode yang dapat diterapkan secara praktis untuk meningkatkan kualitas pembelajaran berbasis video, sehingga lebih menarik dan mudah diakses.

1.6. Kerangka Berpikir

Kerangka berpikir pada Gambar 1.1 di bawah ini menggambarkan alur pendekatan yang digunakan penulis dalam memberikan solusi untuk meningkatkan efektivitas pembelajaran online. Pendekatan ini dirancang untuk mengatasi tantangan dalam memahami materi dengan lebih efisien melalui peringkasan video pembelajaran dan video berita. Berikut adalah kerangka berpikir yang dimaksud:



Gambar 1.1 KerangkaBerpikir

Gambar 1.1 menggambarkan kerangka berpikir dalam penelitian mengenai sistem perangkum teks otomatis dari video pembelajaran dan video berita. Penjelasan dimulai dengan masalah utama, yaitu banyaknya video pembelajaran dan video berita yang memiliki durasi panjang sehingga menyulitkan pengguna untuk memperoleh informasi penting secara cepat. Selain itu, sebagian besar pengguna mungkin kehilangan fokus saat menonton video, yang menyebabkan banyak informasi penting terlewat atau tidak tersampaikan secara efektif.

Untuk mengatasi permasalahan tersebut, peluang solusi yang ditawarkan adalah pemanfaatan teknologi Speech-to-Text (STT) menggunakan Whisper untuk mentranskripsi audio dari video menjadi teks. Setelah teks diperoleh, model Long-T5 berbasis Transformer digunakan untuk meringkas teks transkripsi agar lebih padat dan informatif.

Bagian pendekatan dalam penelitian ini mencakup implementasi sistem yang menggabungkan Whisper untuk transkripsi, dan Long-T5 untuk peringkasan teks. Proses ini memastikan bahwa teks yang dihasilkan tidak hanya akurat tetapi juga diringkas dengan baik untuk memudahkan pemahaman pengguna.

Dalam metode pengembangan, digunakan pendekatan CRISP-DM (Cross-Industry Standard Process for Data Mining) untuk membangun sistem perangkum teks dari video pembelajaran dan video berita. Metode ini dipilih karena struktur kerjanya yang sistematis, mulai dari pemahaman bisnis, pemrosesan data, pemodelan, hingga evaluasi hasil.

Pada tahap implementasi, penelitian ini menggunakan bahasa pemrograman Python serta berbagai library NLP dan Deep Learning untuk melatih model serta mengintegrasikannya ke dalam sebuah platform berbasis web atau aplikasi interaktif.

Akhirnya, dalam hasil yang diharapkan, sistem yang dikembangkan mampu menghasilkan ringkasan otomatis dari video pembelajaran dan video berita dalam bentuk teks. Dengan adanya sistem ini, pengguna dapat memperoleh informasi utama dengan lebih cepat dan efisien, tanpa harus menonton keseluruhan video.

Penjelasan ini memberikan gambaran bagaimana sistem ini dirancang dan diimplementasikan untuk meningkatkan efisiensi dalam memahami materi dari video pembelajaran dan video berita.