

**PODCAST SUMMARIZATION
USING WAV2VEC2 AND BART PRETRAINED MODEL**

**LAPORAN RANCANGAN PROYEK
MATA KULIAH COMP6822001 – SPEECH RECOGNITION
KELAS LA01**



Diusulkan Oleh :

Kania Galih Widowati – 2502047070

Kezia Foejiono - 2540131014

Muhammad Haikal Irawan - 2502000823

Semester Genap Tahun 2023

BINUS UNIVERSITY

DAFTAR ISI

A. Deskripsi Project	3
B. Dataset	3
C. Tahap Eksperimen	3
D. Fitur	4
E. Metode	4
F. Rencana Evaluasi	4

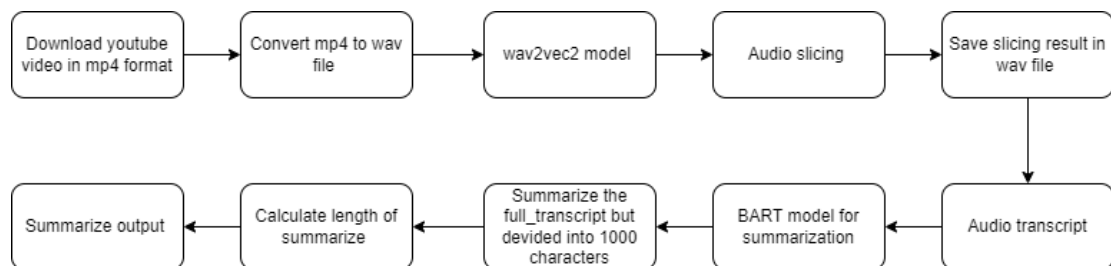
A. Deskripsi Project

Podcast summarization merupakan proses untuk mendapatkan point dan informasi utama dari suatu podcast yang memiliki beragam durasi. Untuk Sebagian orang mendengarkan podcast yang memiliki durasi yang Panjang, bukanlah masalah yang besar. Namun untuk beberapa orang yang memiliki keterbatasan waktu, mendengarkan podcast adalah hal yang merepotkan. Namun dengan terus berkembangnya teknologi dan informasi, permasalahan seperti tidak adanya waktu untuk mendengarkan podcast, bisa teratasi. Tujuan dari pembuatan podcast summarization ini adalah untuk memberikan gambaran singkat kepada pendengar. Hal ini berguna untuk memahami ide utama dari podcast yang disampaikan tanpa harus mendengar seluruh episode. Hal ini juga bermanfaat untuk pengguna dimana mereka dapat dengan mudah memutuskan apakah episode yang akan mereka tonton ini sepadan dengan waktu yang akan mereka luangkan untuk mendapatkan informasi yang mereka inginkan. Pada model yang kami buat ini memiliki beberapa Langkah yang akan dijelaskan pada bagian Tahap Eksperimen. Namun secara gambaran umum terdapat dua Langkah utama yaitu transcript dan summarization. Pada Audio transcript kami menggunakan pretrained model wav2vec2 dan pada summarization menggunakan BART pretrained model.

B. Dataset

Pada pretrained BART model menggunakan CNN Daily Mail Dataset.

C. Tahap Eksperimen



Dalam membangun Podcast Summarization Model terdapat beberapa tahapan :

1. Download video dari youtube, hal pertama yang perlu dilakukan adalah melakukan install library pytub yang digunakan untuk mendownload video youtube. Kemudian membuat variable dengan nama youtube_video yang berisikan link youtube yang akan di download. Kemudian melakukan filtering untuk mendownload format audion MP4 kemudian mendownload file.
2. Melakukan conversi file dengan format MP4 menjadi wav.
3. Pada bagian Automatic Speech Recognition, install torch dan huggingsound. Kemudian melakukan pengecekan apakah GPU terdapat cuda. Jika ya maka nantinya akan menggunakan cuda, jika pada GPU tidak terdapat CUDA maka akan menggunakan CPU. Setelah itu panggil pretrained model wav2vec2.
4. Audio slicing, melakukan slicing dengan durasi 30 detik dan menetapkan audio frame = 16000. Kemudian setelah melakukan slicing, simpan file dengan format wav.

5. Melakukan Audio Transcript menggunakan model wav2vec, kemudian menghitung Panjang character yang dihasilkan dari proses transcript.
6. Pada bagian text summarization, Langkah awal yang perlu dilakukan adalah install transformer dan import pipeline karena akan menggunakan function dari library pipeline. Model yang digunakan adalah BART. Kemudian melakukan summarization.

D. Fitur

Pada model Podcast Summarization yang kami buat memiliki dua Fitur yaitu **Audio transcript** dan **text summarization**.

- Audio Transcription merupakan proses mengubah sinyal suara kedalam format teks. Pada fitur ini kami menggunakan Automatic Speech Recognition (ASR) menggunakan pretrained Wav2vec2. Audio Transcription sangat penting untuk dilakukan karena akan mempermudah proses pada Text Summarization.
- Fitur kedua yaitu Text Summarization, merupakan proses untuk mengambil point dan informasi utama pada sebuah teks yang dihasilkan dari Audio Transcription. Kami menggunakan Extractive summarization sehingga akan menghasilkan ringkasan yang relevan dan informatif. Hal ini dilakukan dengan pemilihan serta penggabungan kalimat penting yang diambil dari hasil transkripsi.

E. Metode

Pada podcast summarization ini menggunakan dua method. Pada Automatic Speech Recognition menggunakan method wav2vec2 dan pada bagian text summarization menggunakan BART.

F. Rencana Evaluasi

Pada wav2vec2 kami menggunakan Word Error Rate (WER). Sedangkan pada BART kami menggunakan (Recall Oriented Understudy for Gisting Evaluation) ROUGE.

Hasil ROUGE :

	Rouge 1	Rouge 2	Rouge L
r	0.4768392370572207	0.32100840336134456	0.44959128065395093
p	0.5555555555555556	0.3315972222222222	0.5238095238095238
F1-score	0.513196475967484	0.32621690362642314	0.483870962771003