

# **BAB 1**

## **PENDAHULUAN**

### **1.1 Latar Belakang Masalah**

Bumi sebagai tempat tinggal miliaran manusia dan terbentang sangat luas bukan lagi menjadi penghalang untuk bertukar informasi dari satu tempat ke tempat yang lain. Dengan kemajuan teknologi dan lahirnya media elektronik seperti televisi, radio, dan internet, semua informasi dapat tersebar bahkan ke belahan dunia lainnya dengan sangat cepat. Informasi yang berisi suatu peristiwa di suatu tempat dan disebar ke masyarakat luas biasanya dikenal sebagai berita. Secara harfiah, berita adalah suatu kejadian atau peristiwa penting yang terjadi di suatu tempat dan dipublikasikan kepada masyarakat luas sehingga masyarakat yang jauh dari peristiwa tersebut dapat mengetahui apa yang sedang terjadi [1]. Berita merupakan media surat kabar yang sangat penting untuk menyebarluaskan suatu kejadian penting yang baru terjadi di suatu tempat. Secara singkat, berita merupakan hasil dari kegiatan panjang yang disebut jurnalistik yang merupakan segala jenis kegiatan yang menyangkut kegiatan kewartawanan dan persuratkabaran yang di dalamnya terdapat beberapa proses seperti peliputan bahan berita, pelaporan peristiwa, penulisan berita, penyuntingan naskah berita, dan penyajian atau penyebarluasan berita melalui media (Lampiran 1, A-01).

Semua tahap persiapan hingga penyajian berita ke masyarakat terlihat normal jika dilihat oleh masyarakat awam. Namun jika dilihat dari kacamata penggelut dunia jurnalistik khususnya para wartawan, muncul beberapa masalah ketidakefisienan. Salah satu permasalahannya ada pada tahap dokumentasi hasil peliputan berita. Hasil peliputan berita yang dilakukan oleh wartawan biasanya berupa dokumen audio atau video digital yang berisi rekaman berlangsungnya proses wawancara. Proses pengolahan rekaman menjadi transkrip hasil wawancara, masih banyak dilakukan secara manual oleh para wartawan dengan mendengarkan ulang rekaman wawancara dan kemudian menuliskan dokumentasi transkrip hasil wawancara untuk dilanjutkan ke proses berikutnya (Lampiran 1, A-02).

Melalui hasil penelitian kualitatif dengan pakar di bidang jurnalistik, didapat kesimpulan bahwa kebanyakan wartawan di Indonesia masih kurang efisien dari segi waktu ketika membuat transkrip hasil wawancara. Proses tersebut masih dilakukan dengan cara mendengarkan rekaman audio berulang-ulang untuk

menghasilkan transkrip hasil wawancara yang baik (Lampiran 1, A-02). Selain itu, disampaikan juga bahwa tantangan utama dalam pembuatan transkrip hasil wawancara ini adalah kesesuaian hasil transkrip dengan wawancara yang dilakukan dan kecepatan dalam pembuatannya. Meskipun dengan cara tradisional hasil transkripsi yang dihasilkan mempunyai kualitas yang baik, namun waktu yang dialokasikan untuk membuatnya sangat kurang efisien (Lampiran 1, A-03).

Permasalahan yang disampaikan oleh pakar di bidang jurnalistik sebelumnya sejalan dengan adanya teknologi *Automatic Speech Recognition (ASR)*. Teknologi *ASR* memungkinkan mesin untuk memahami ucapan pengguna dan mengubahnya menjadi rangkaian kata melalui program komputer, sehingga menciptakan semacam komunikasi alami antara manusia dan mesin [2]. Teknologi *ASR* ini mempunyai tujuan utama untuk mentranslasi suara atau bahasa lisan menjadi teks secara otomatis [3].

Penelitian terkait pembangunan sistem *ASR* untuk bahasa Indonesia telah dilakukan oleh beberapa peneliti terdahulu. Salah satunya adalah penelitian dengan menggunakan *Unidirectional Gated Recurrent Unit* yang menunjukkan hasil kurang baik dengan metrik pengukuran *Word Error Rate (WER)* bernilai 90.61% [4]. Penelitian lain juga dilakukan menggunakan *Bidirectional Long Short-Term Memory* dan *Connectionist Temporal Classification* untuk melatih model yang menerjemahkan suara bahasa Indonesia menjadi teks. Hasil penelitian menunjukkan performa yang lebih baik dari penelitian sebelumnya yaitu dengan *WER* yaitu 20,87% [5].

Di luar beberapa penelitian terdahulu, terdapat penelitian terkini terkait model *Wav2Vec2* yang dikembangkan oleh tim *AI Facebook* yang dikenal sebagai solusi *ASR* untuk *Low-Resource Language*. *Low-resource language* merupakan istilah dalam *Machine Learning* yang mengacu pada bahasa yang memiliki data yang relatif lebih sedikit untuk melatih sistem AI percakapan [6]. Hasil penelitian menggunakan *Wav2Vec2* menunjukkan hasil yang baik dengan hasil *Word Error Rate (WER)* rata-rata adalah 28,83% untuk 6 bahasa berbeda, sedangkan hasil yang paling optimal ditunjukkan oleh bahasa Inggris yaitu 16,07% *WER* dengan menggunakan total 15 jam data latih suara [7].

Penelitian mendalam terkait penggunaan model *Wav2Vec2* untuk bahasa Indonesia masih belum dilakukan, baik dari segi konfigurasi optimal dan hasil performanya. Oleh karena itu, peneliti bermaksud untuk melakukan penelitian terkait model *Wav2Vec2* sebagai solusi *ASR* bahasa Indonesia. Hasil dari model pengenalan suara yang telah dilatih akan diimplementasikan menjadi sistem

aplikasi transkripsi suara sehingga diharapkan dapat digunakan dan membantu menyelesaikan permasalahan di bidang jurnalistik terutama dalam hal transkripsi hasil wawancara.

## **1.2 Rumusan Masalah**

1. Bagaimana implementasi model Wav2Vec2 untuk sistem transkripsi percakapan suara bahasa Indonesia dalam kegiatan jurnalistik, khususnya dalam transkripsi hasil wawancara?
2. Bagaimana hasil akurasi dan waktu transkripsi model pengenalan suara bahasa sumber daya rendah yang dilatih dengan data suara bahasa Indonesia dengan metrik pengukuran *Word Error Rate* dan *Evaluation Metrics*?

## **1.3 Batasan Permasalahan**

1. Sistem transkripsi percakapan suara terbatas secara 1 arah yaitu dari percakapan suara menjadi teks dan tidak sebaliknya.
2. Data latih yang digunakan untuk melatih model pengenalan suara terbatas pada data bahasa Indonesia dengan jumlah 5809 audio suara yang didapat secara terbuka dari bank data *commonvoice* dengan rasio data pelatihan dan pengujian adalah 8:2.
3. Model pengenalan suara dilatih dengan terbatas pada huruf alphabet bahasa Indonesia tanpa memperhatikan tanda baca atau simbol lainnya.

## **1.4 Tujuan Penelitian**

1. Membangun sistem transkripsi percakapan suara bahasa Indonesia dengan model ASR Wav2Vec2 yang diharapkan dapat bermanfaat terutama dalam kegiatan jurnalistik, khususnya dalam proses transkripsi hasil wawancara.
2. Mengukur dan mengetahui tingkat akurasi dan efektivitas waktu transkripsi percakapan suara bahasa Indonesia yang telah dikembangkan dengan metrik pengukuran *Word Error Rate*.

## 1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat bagi pihak-pihak yang terkait dengan penelitian ini, antara lain:

1. Bagi Peneliti.  
Memperluas ilmu dan pengetahuan dalam bidang kecerdasan buatan terutama dalam pengenalan suara bahasa Indonesia, serta sebagai salah satu syarat untuk meraih gelar Sarjana Strata Satu (S1) di Universitas Multimedia Nusantara.
2. Bagi Bidang Jurnalistik.  
Memudahkan pekerjaan para jurnalis, wartawan, dan reporter dalam mengumpulkan informasi atau berita, terutama dalam proses transkripsi percakapan dengan narasumber.
3. Bagi Masyarakat Umum.  
Memudahkan dalam melakukan transkripsi suara bahasa Indonesia menjadi teks untuk segala macam kebutuhan.

## 1.6 Sistematika Penulisan

Berisikan uraian singkat mengenai struktur isi penulisan laporan penelitian, dimulai dari Pendahuluan hingga Simpulan dan Saran.

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN  
Bab pertama memaparkan latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan laporan.
- Bab 2 LANDASAN TEORI  
Bab kedua memaparkan teori-teori yang mendasari penelitian ini, antara lain jurnalistik, wawancara, transkripsi percakapan, *Automatic Speech Recognition*, *Data Pre Processing*, *Fine Tuning*, *Wav2Vec2*, dan metrik evaluasi *Word Error Rate*.
- Bab 3 METODOLOGI PENELITIAN  
Bab ketiga memaparkan tahapan yang akan dilakukan dari sebelum sampai proses penelitian selesai. Pada bab ini juga dilengkapi dengan flowchart dan diagram.

- Bab 4 HASIL DAN DISKUSI

Bab keempat memaparkan implementasi kode, skenario yang diuji, hasil pengujian, serta evaluasi terhadap hasil yang diperoleh.

- Bab 5 KESIMPULAN DAN SARAN

Bab kelima memaparkan kesimpulan yang menjawab tujuan penelitian dan saran untuk pengembangan ataupun penelitian yang bersangkutan di kemudian hari.

