

BAB I

PENDAHULUAN

1.1 Latar Belakang

Emosi merupakan komponen penting dalam komunikasi manusia karena tidak hanya menyampaikan informasi faktual, tetapi juga memperkuat makna dan intensi dari suatu ucapan. Dalam pembelajaran bahasa asing, khususnya bahasa Inggris, banyak penutur non-native mengalami kesulitan dalam mengekspresikan emosi secara tepat karena keterbatasan dalam memahami prosodi seperti intonasi, tekanan, dan ritme. Hal ini menyebabkan ekspresi emosional melalui suara menjadi kurang akurat atau bahkan salah makna.

Permasalahan yang sering muncul adalah kurangnya kemampuan penutur dalam memahami atau menyampaikan emosi saat berbicara dalam bahasa Inggris, yang dapat menghambat efektivitas komunikasi, terutama dalam konteks pembelajaran atau layanan interaktif berbasis suara. Selain itu, tidak semua aplikasi mampu menangkap makna emosional dari sebuah ujaran jika hanya mengandalkan pengenalan suara tanpa memperhatikan struktur dan makna bahasa.

Untuk mengatasi masalah ini, diperlukan sebuah solusi teknologi yang dapat membantu pembelajar bahasa Inggris memahami emosi secara lebih eksplisit tanpa hanya bergantung pada fitur prosodi yang kompleks. Urgensi dari penelitian ini terletak pada perlunya pendekatan baru yang mampu mendeteksi emosi tanpa harus mengandalkan fitur prosodik seperti nada dan intonasi, melainkan melalui analisis struktur bahasa dan makna kata dari hasil transkripsi suara. Hal ini membuka peluang bagi pengembangan teknologi interaksi manusia-komputer yang lebih inklusif, terutama bagi mereka yang belajar bahasa asing atau memiliki keterbatasan dalam mengungkapkan emosi melalui intonasi. Dengan memanfaatkan teknologi *Speech-to-Text* (STT) untuk mengubah suara menjadi teks dan *Natural Language Processing* (NLP) untuk menganalisis teks tersebut, aplikasi ini menawarkan alat bantu praktis. Sistem ini akan memberikan umpan balik langsung kepada pengguna mengenai emosi yang terkandung dalam ucapan mereka, memungkinkan mereka untuk melatih dan meningkatkan pemahaman

emosional dalam komunikasi verbal berbahasa Inggris. Dengan demikian, penelitian ini tidak hanya berkontribusi pada pengembangan teknologi interaksi manusia-komputer, tetapi juga menyediakan solusi edukatif yang relevan untuk mengatasi tantangan nyata dalam pembelajaran bahasa asing. Penelitian sebelumnya telah menunjukkan bahwa banyak ekspresi emosi seperti marah, senang, atau sedih dapat dikenali dari konteks bahasa itu sendiri (Amalia Khaerunnisa Nursya Bany, 2022), bahkan dalam format teks yang telah diproses. Dengan demikian, meskipun tidak menggunakan fitur *prosodi*, pendekatan berbasis teks tetap efektif terutama dalam konteks edukatif untuk membantu pembelajar bahasa Inggris mengenali dan memahami emosi dalam komunikasi verbal.

Penelitian ini bertujuan untuk mengembangkan aplikasi web yang dapat mendeteksi emosi dari suara yang sudah di *transkrip* menjadi text menggunakan pendekatan *Speech-to-Text* (STT) dan *Natural Language Processing* (NLP). Aplikasi ini tidak menganalisis nada suara, ritme bicara, atau pitch (*prosodi*), melainkan hanya berfokus pada teks hasil *transkripsi* suara sebagai dasar pendeteksian emosi. Dengan demikian, pengguna cukup mengucapkan kalimat dalam bahasa Inggris, lalu sistem akan mengonversinya menjadi teks dan menganalisis emosi dari struktur dan pilihan kata dalam kalimat tersebut.

Speech-to-Text (STT) merupakan teknologi yang mengubah sinyal suara menjadi representasi teks. STT telah banyak digunakan dalam berbagai aplikasi seperti asisten virtual, sistem navigasi, dan layanan pelanggan otomatis. Salah satu metrik untuk menilai performa STT adalah *Word Error Rate* (WER), yaitu ukuran seberapa banyak kesalahan dalam hasil transkripsi dibandingkan dengan transkripsi referensi. Semakin rendah nilai WER, semakin baik kualitas transkripsi yang dihasilkan.

Setelah suara dikonversi menjadi teks, proses deteksi emosi dilakukan menggunakan pendekatan *Natural Language Processing* (NLP). NLP merupakan cabang dari kecerdasan buatan yang memungkinkan sistem untuk memahami, menafsirkan, dan menganalisis bahasa alami yang digunakan oleh manusia. Dalam konteks penelitian ini, NLP digunakan untuk mengidentifikasi emosi berdasarkan

pola linguistik dalam kalimat, seperti penggunaan kata-kata berkonotasi emosional, struktur kalimat, dan konteks lokal. Meskipun pendekatan berbasis teks tidak dapat menggantikan informasi prosodik seperti intonasi atau pitch yang sering kali menjadi indikator kuat emosi, penelitian-penelitian sebelumnya telah menunjukkan bahwa teks tetap dapat merepresentasikan emosi dengan akurasi yang cukup tinggi, terutama dalam konteks kalimat yang eksplisit (van Schijndel, Nguyen, & Schuler, 2013).

Penelitian dari (Fauzul, Kustiwa, Aminul Akbar, & Pinandito, 2017) menunjukkan bahwa metode *leksikon* dan NLP mampu mengklasifikasikan emosi dari teks media sosial dengan cukup akurat. (Fauzul et al., 2017) dalam studinya mengoptimalkan model *RoBERTa* untuk mendeteksi emosi dari teks berbahasa Indonesia dan mencapai akurasi hingga 81%. Meski demikian, kedua penelitian ini tidak membahas bagaimana emosi bisa dideteksi dari hasil transkripsi suara, dan lebih menitikberatkan pada teks tertulis.

Dengan menggabungkan teknologi STT dan NLP ke dalam sebuah aplikasi berbasis web, penelitian ini menawarkan pendekatan baru yang dapat diakses secara luas dan mudah digunakan. Sistem ini dirancang untuk membantu pengguna khususnya pembelajar bahasa Inggris dalam melatih kemampuan memahami emosi dari ucapan yang mereka hasilkan sendiri, tanpa harus bergantung pada kemampuan mereka menangkap atau menyampaikan nada dan intonasi secara tepat. Pendekatan ini sekaligus menjembatani kesenjangan antara aspek linguistik dan ekspresi emosional dalam proses belajar bahasa asing.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah yang akan dibahas dalam penelitian ini adalah :

1. Bagaimana cara mengembangkan aplikasi web yang mampu mengonversi suara menjadi teks secara akurat menggunakan metode *Speech-to-Text*?
2. Bagaimana cara menggunakan metode NLP untuk mendeteksi emosi dari teks hasil pengenalan suara?
3. Bagaimana cara mengevaluasi 2 metode STT dan NLP secara objektif?

1.3 Tujuan Penelitian

Adapun penelitian ini memiliki tujuan, diantaranya adalah :

1. Mengembangkan aplikasi web yang mampu mengenali suara dan mengonversinya menjadi teks dengan metode *Speech-to-Text*.
2. Menerapkan metode NLP untuk menganalisis dan mendeteksi emosi dari teks hasil pengenalan suara.
3. Mengevaluasi tingkat akurasi sistem dalam mengenali dan mendeteksi emosi menggunakan metrik evaluasi seperti *Word Error Rate* (WER) untuk *Speech-to-Text* dan *Confusion Matrix* untuk NLP.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat membawa manfaat. Adapun manfaat dari penelitian ini adalah :

1. Memberikan kontribusi dalam pengembangan teknologi interaksi manusia dan komputer yang lebih manusiawi, khususnya melalui suara.
2. Menjadi referensi bagi pengembangan lebih lanjut di bidang pengenalan suara dan deteksi emosi dengan integrasi teknologi STT dan NLP.

1.5 Ruang Lingkup dan Batasan Masalah

Untuk menjaga fokus dan keterukuran penelitian, ruang lingkup dan batasan masalah yang diterapkan adalah sebagai berikut:

1. Bahasa yang didukung terbatas pada Bahasa Inggris standar. Sistem hanya menerima masukan suara dalam Bahasa Inggris dan tidak memperhitungkan aksen atau dialek lokal secara khusus.
2. Jenis emosi yang dideteksi dibatasi pada empat kategori utama, yaitu senang (happy), sedih (sad), marah (Aggry), dan netral (neutral).
3. Dataset yang digunakan adalah dataset publik dari *Common Voice by Mozilla*. Dataset ini digunakan untuk pengujian model dalam konteks transkripsi dan analisis emosi berbasis text sebanyak 320 rekaman.
4. Penelitian ini tidak mencakup pengenalan percakapan dua arah secara *real-time*.

sistem ini berfokus pada analisis emosi yang bersumber dari makna bahasa (*textual content*) hasil konversi suara ke teks. Penelitian ini tidak menganalisis fitur-fitur akustik seperti nada (*pitch*), intonasi, atau ritme suara (*prosody*). Penekanan diberikan pada kemampuan model *Lexicon-based* dalam mengidentifikasi bobot emosi dari struktur kalimat yang dihasilkan oleh teknologi *Speech-to-Text*.