

BAB I

PENDAHULUAN

1.1 Latar Belakang

Interaksi antara satu individu dengan yang lainnya sering terjadi dengan cara berbicara secara lisan atau berdialog. Pembicaraan atau dialog tersebut terdapat beberapa faktor seperti tinggi rendah intonasi, kalimat yang diucap, dan maksud dari pembicara. Hal-hal tersebut berpengaruh dalam emosi seseorang saat mengucapkan suatu kalimat. *Speech Recognition* adalah bidang yang mendeteksi emosi dari sinyal suara. Beberapa penelitian memperlihatkan ada hubungan antara satuan statistik ucapan dan emosi dari pembicara. Satuan-satuan yang sering digunakan diantaranya adalah energi, nada, artikulasi, dan bentuk spektral. Satuan-satuan tersebut membentuk sebuah ciri dari suatu emosi seperti marah, sedih, bahagia, terkejut, dan lain-lain. Ciri tersebut dapat diklasifikasi dengan beberapa metode yang pernah diujikan sebelumnya[1].

Beberapa metode klasifikasi emosi yang pernah digunakan pada penelitian sebelumnya adalah *Mel Frequency Cepstral Coefficient (MFCC)* dan *K-Nearest Neighbour (K-NN)* dengan dataset sebanyak 200 suara musik mendapatkan akurasi 85.5%[2]. Penelitian lain menggunakan metode *Zero Crossing Rate (ZCR)* dan *Convolutional Neural Network (CNN)* dengan dataset sebanyak 1231 rekaman suara yang mendapatkan akurasi sebesar 48.05% [3].

Metode yang banyak digunakan dalam penelitian ilmiah adalah *Mel Frequency Cepstral Coefficient (MFCC)*, yang digunakan untuk mengekstraksi ciri-ciri dari sinyal audio. *MFCC* telah menjadi pendekatan utama dalam berbagai bidang pemrosesan suara, merupakan teknik terkemuka dalam ekstraksi fitur ucapan, dan terbukti efektif dalam menghasilkan ciri-ciri yang dapat digunakan untuk klasifikasi emosi[1][4]. Dasar dari *MFCC* adalah konsep variasi *bandwidth* kritis terhadap frekuensi yang dapat dideteksi oleh pendengaran manusia.

Zero Crossing Rate (ZCR) adalah salah satu metode ekstraksi ciri dari sinyal audio yang berfokus pada domain waktu. *ZCR* dapat dijelaskan sebagai hitungan

seberapa sering dalam suatu periode waktu tertentu atau frame, amplitudo sinyal audio melewati nilai nol [3].

Berdasarkan uraian diatas, peneliti ingin mendapatkan akurasi dari mengambil 2 jenis fitur dari audio suara yaitu *MFCC* dan *ZCR* lalu menggunakan klasifikasi *K-NN*. Dataset yang dipilih untuk penelitian ini merupakan 1440 data suara manusia yang didapat dari *Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)*.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Banyak penelitian mengenai klasifikasi emosi yang telah mengadopsi penerapan *machine learning* sebagai media pembelajaran. Namun pada beberapa penelitian yang sudah ada, penggunaan *machine learning* untuk klasifikasi emosi masih menggunakan teks sebagai masukannya. Penelitian ini bertujuan untuk mengembangkan sebuah model analisis emosi pada data audio menggunakan fitur *Mel Frequency Cepstral Coefficients (MFCC)* dan *Zero Crossing Rate (ZCR)* dengan algoritma *K-Nearest Neighbour (KNN)*.

1.2.2 Metode Usulan

Untuk menjawab permasalahan tersebut, metode usulan yang digunakan adalah pengolahan data audio menggunakan fitur *MFCC* dan *ZCR* serta algoritma *KNN*. Metode ini telah digunakan dalam beberapa penelitian sebelumnya dan terbukti efektif dalam klasifikasi emosi pada data audio. Data audio yang digunakan berasal dari dataset *RAVDESS* [6]. Selanjutnya, fitur-fitur tersebut digunakan sebagai input untuk algoritma *KNN* dalam proses klasifikasi emosi.

1.2.3 Pertanyaan Penelitian

1. berapa akurasi yang diperoleh dari menggunakan fitur *MFCC* dengan klasifikasi *K-NN* ?
2. berapa akurasi yang diperoleh dari menggunakan fitur *ZCR* dengan klasifikasi *K-NN* ?

3. berapa akurasi yang diperoleh dari menggunakan fitur *MFCC* dan *ZCR* dengan klasifikasi *K-NN* ?

1.3 Tujuan dan Manfaat

1.2.1 Tujuan

Mengukur tingkat akurasi dari penerapan metode *K-Nearest Neighbour* dalam klasifikasi emosi dari fitur *MFCC* dan *ZCR* yang diperoleh dari signal audio, serta mencari classifier *KNN* yang terbaik.

1.2.2 Manfaat

Dapat membantu dalam pengklasifikasian emosi melalui signal audio suara manusia.

1.4 Batasan-batasan

Batasan masalah dalam penelitian ini meliputi:

1. Penelitian ini hanya difokuskan pada klasifikasi emosi manusia melalui data *speech audio*.
2. Dataset yang digunakan adalah dataset dari *RAVDESS* yang berisikan 1440 data audio wicara.
3. Fitur yang digunakan dalam penelitian ini terbatas pada *MFCC* dan *Zero Crossing Rate*.
4. Algoritma yang digunakan untuk klasifikasi emosi adalah *K-Nearest Neighbor (KNN)*.
5. Bahasa Pemrograman yang digunakan adalah *Python*.
6. Ekstraksi Fitur menggunakan *Library Librosa*.
7. Klasifikasi emosi yang digunakan ada 8 emosi
8. Penelitian ini tidak mempertimbangkan faktor latar belakang budaya dan sosial dari pembicara dalam data audio.

1.5 Sistematika Skripsi

Proses penulisan skripsi ini dibagi menjadi beberapa beberapa bab.

BAB I : PENDAHULUAN

Pendahuluan berisi latar belakang, rumusan masalah, uraian masalah, tujuan dan manfaat, dan sistematika laporan, menjelaskan dasar-dasar pembuatan tugas akhir.

BAB II : KAJIAN PUSTAKA

Tinjauan masalah berbicara tentang penelitian sebelumnya yang telah dilakukan.

BAB III : METODE USULAN

Metodologi penelitian berbicara bagaimana alur sistem dibuat pada penelitian ini, seperti arsitektur sistem, preprocessing, pembobotan kata, skenario pengujian

BAB IV : HASIL DAN PEMBAHASAN

Berisi hasil dari penerapan metode penelitian serta penjelasan hasil penelitian, *source code*, dan *user interface website*.

BAB V : PENUTUP

Berisi kesimpulan dari penelitian yang telah dilakukan dan saran untuk penelitian selanjutnya