

BAB I

PENDAHULUAN

1.1. Latar Belakang

Ucapan adalah metode yang disukai dan paling sering digunakan oleh manusia untuk berkomunikasi antara satu sama lain. Karena dengan ucapan manusia dapat membagikan informasi dan perasaannya dengan lebih cepat (Jain, et al., 2018). Ucapan itu sendiri berisi informasi yang bervariasi dimana selain dapat mengetahui pesan dan maksud seseorang, kita juga dapat mengetahui kondisi emosi orang tersebut (Helmiyah, et al., 2020). Emosi sendiri merupakan bentuk ekspresi manusia terhadap suatu hal ataupun seseorang. Ekspresi emosi dalam sebuah percakapan berperan penting dalam memberikan penekanan pada informasi yang disampaikan menjadi lebih kuat daripada hanya dengan kata-kata saja (Lasiman & Lestari, 2018). Hal ini menyebabkan penelitian untuk mendeteksi emosi manusia melalui sebuah ucapan sedang giat dilakukan.

Pengenalan emosi melalui sebuah ucapan adalah tindakan mencoba mengenali emosi manusia dan keadaan afektif dari sebuah percakapan. Hal ini didasarkan pada fakta bahwa setiap suara yang dikeluarkan mencerminkan emosi melalui nada dan intonasi. Pengenalan emosi melalui ucapan ini sangat berguna dan dapat diaplikasikan ke dalam berbagai bidang seperti ilmu kognitif, psikologi, *neuroscience*, *call centre*, industri game, kesehatan dan bidang lainnya (Jain, et al., 2018). Salah satu contoh penerapannya adalah pada kokpit pesawat, dimana digunakan untuk mengetahui kondisi mental dari pilot sehingga dapat menghindari terjadinya kecelakaan. Selain itu juga pada percakapan *call centre*, untuk dapat mengetahui kondisi mood seorang pelanggan sehingga dapat digunakan untuk meningkatkan kualitas pelayanan *call centre* tersebut.

Pesatnya perkembangan teknologi informasi saat ini yang mana tidak lepas mengenai bidang pemrosesan sinyal digital. Memungkinkan perancangan suatu sistem cerdas yang dapat mengidentifikasi jenis emosi manusia melalui ucapan secara otomatis. Dalam kasus ini, penulis akan membuat sistem yang dapat

mengidentifikasi jenis emosi manusia berdasarkan ucapan dengan menggunakan metode *Convolutional Neural Network*.

Berdasarkan studi literatur yang dilakukan oleh penulis terbukti metode *convolutional neural network* ini cocok dan memberikan hasil akurasi yang baik dalam memprediksi kelas emosi seseorang dikarenakan adanya kompleksitas pada fiturnya (Jain, et al., 2018). Penelitian yang serupa dilakukan oleh (Zhao, Mao, & Chen, 2019; Qayyum, Arefeen, & Shahnaz, 2019) dimana didapatkan hasil penggunaan dengan metode CNN untuk identifikasi emosi memberikan hasil akurasi yang lebih tinggi dibandingkan dengan menggunakan metode lain seperti *Support Vector Machine* (SVM), *Multivariate Linear Regression Classification* (MLR) dan *Recurrent Neural Network* (RNN). Dari segi ekstraksi fitur pada suara, salah satu metode utama yang digunakan oleh penulis pada penelitian ini adalah *Mel Frequency Cepstrum Coefficient* (MFCC), MFCC merupakan metode yang bekerja dengan meniru cara telinga manusia bekerja dalam membedakan tinggi rendahnya suara, sehingga dapat mengetahui ciri dari suara tersebut. Setiap audio percakapan sebelumnya akan dilakukan ekstraksi fitur terlebih dahulu, sebelum nantinya dilakukan pengklasifikasian menggunakan metode *Convolutional Neural Network*.

Pada penelitian sebelumnya dengan judul "Pemodelan CNN Untuk Deteksi Emosi Berbasis Speech Bahasa Indonesia", memiliki kekurangan dimana hanya menggunakan dataset yang berukuran kecil serta jenis emosi yang hanya berjumlah empat yaitu marah, senang, netral dan sedih. Dari permasalahan tersebut pada penelitian kali ini penulis juga memutuskan untuk menggunakan dataset yang berukuran lebih besar serta jenis kelas emosi yang lebih banyak daripada penelitian sebelumnya. Pada penelitian kali ini penulis menggunakan dataset *Ryerson Audio-Visual Database of Emotional Speech and Song* (RAVDESS) yang sangat terkenal pada bidang *Speech Emotion Recognition* (SER), dataset ini sendiri terdiri dari 8 jenis emosi yaitu senang, sedih, marah, takut, terkejut, jijik, tenang dan netral. Setiap kelas emosi masing-masing memiliki 192 *file audio* jadi total keseluruhan dataset ini ada 1440 *file audio*. Dengan menggunakan dataset yang lebih besar dan jenis emosi yang lebih banyak maka model akan dapat mengenali lebih banyak jenis emosi manusia.

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dipaparkan, maka rumusan masalah adalah sebagai berikut :

1. Bagaimana penerapan metode *Convolutional Neural Network* untuk mendeteksi 8 kelas emosi melalui ucapan?
2. Bagaimana hasil akurasi model sesudah menggunakan metode *Convolutional Neural Network* untuk deteksi kelas emosi melalui ucapan?

1.3. Tujuan

Tujuan penelitian adalah hasil yang ingin dicapai dalam penelitian, maka tujuan penelitian ini adalah :

1. Untuk mengetahui bagaimana proses mendeteksi 8 kelas emosi melalui ucapan menggunakan metode *Convolutional Neural Network*.
2. Mencari hasil akurasi yang lebih baik dalam mendeteksi kelas emosi dengan menggunakan metode *Convolutional Neural Network*.

1.4. Manfaat

Beberapa manfaat yang didapatkan dari penelitian ini antara lain sebagai berikut :

1. Bagi penulis
Mendapatkan pengetahuan mengenai bagaimana cara mendeteksi sebuah emosi melalui ucapan manusia menggunakan pembelajaran mesin. Setelah penelitian ini penulis mengetahui metode yang tepat serta langkah-langkah penyelesaian untuk dapat mendeteksi emosi melalui ucapan manusia.
2. Bagi peneliti lain
Hasil penelitian ini dapat dijadikan referensi dalam bidang kecerdasan buatan untuk penelitian-penelitian selanjutnya yang lebih kompleks.

3. Bagi instansi terkait

Hasil penelitian ini diharapkan dapat memberikan manfaat bagi pihak yang berkepentingan seperti dalam bidang psikologi dapat diterapkan untuk mendeteksi emosi melalui ucapan, sehingga psikiater dapat bekerja dengan lebih efisien, dan juga bagi pembaca dapat membedakan emosi berdasarkan jenisnya.

1.5. Batasan Masalah

Batasan masalah diperlukan dalam sebuah penelitian agar penelitian tersebut tidak menyimpang dan fokus pada masalah yang akan dibahas. Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Data pada penelitian ini diambil dari *dataset* pada kaggle yaitu RAVDESS yang terdiri dari 1440 file suara. File suara pada dataset ini terdiri dari suara laki-laki dan perempuan yang masing-masing berjumlah 12 orang.
2. Sampel suara yang digunakan pada dataset masih dalam bahasa inggris.
3. Setiap sampel suara pada dataset hanya memiliki satu kelas emosi.
4. Terdapat 8 kelas emosi yang dapat dikenali oleh model yang akan dibuat yaitu, emosi netral, tenang, sedih, senang, takut, marah, jijik dan terkejut.