ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v20i1.9518

Implementation of Mel-Frequency Cepstral Coefficient as Feature Extraction using K-Nearest Neighbor for Emotion Detection Based on Voice Intonation

Implementasi Ekstraksi Ciri Mel-Frequency Cepstral Coefficient Menggunakan K-Nearest Neighbor Untuk Deteksi Emosi Berdasarkan Intonasi Suara

Revanto Alif Nawasta[1], Nur Heri Cahyana[2], Heriyanto[3]

[1,2,3] Informatika, Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia

[1*] 123170060@student.upnyk.ac.id, [2]nur.hericahyana@upnyk.ac.id, [3]heriyanto@upnyk.ac.id
**: Penulis korenspondensi* (*corresponding author*)

Informasi Artikel
Received: December 2022
Revised: January 2023
Accepted: January 2023
Published: February 2023

Abstract

Purpose: To determine emotions based on voice intonation by implementing MFCC as a feature extraction method and KNN as an emotion detection method.
Design/methodology/approach: In this study, the data used was downloaded from several video podcasts on YouTube. Some of the methods used in this study are pitch shifting for data augmentation, MFCC for feature extraction on audio data, basic statistics for taking the mean, median, min, max, standard deviation for each coefficient, Min max scaler for the normalization process and KNN for the method classification.
Findings/result: Because testing is carried out separately for each gender, there are two classification models. In the male model, the highest accuracy was obtained at 88.8% and is included in the good fit model. In the female model, the highest accuracy was obtained at 92.5%, but the model was unable to correctly classify emotions in the new data. This condition is called overfitting. After testing, the cause of this condition was because the pitch shifting augmentation process of one tone in women was unable to solve the problem of the training data size being too small and not containing enough data samples to accurately represent all possible input data values.
Originality/value/state of the art: The research data used in this study has never been used in previous studies because the research data is obtained by downloading from Youtube and then processed until the data is ready to be used for research.

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v20i1.9518

Keywords: Mel-Frequency Cepstral Coefficient; K-Nearest Neighbor; Emotion detection; Signal processing
Kata kunci: Deteksi emosi; K-Nearest Neighbor; Mel-Frequency Cepstral Coefficient; Pemrosesan Sinyal

Abstrak

Tujuan: Untuk mengetahui emosi berdasarkan intonasi suara dengan mengimplementasikan MFCC sebagai metode ekstraksi ciri dan KNN sebagai metode deteksi emosi.
Perancangan/metode/pendekatan: Pada penelitian ini data yang digunakan diunduh dari beberapa video *podcast* di *youtube*. Beberapa metode yang digunakan pada penelitian ini yaitu *pitch shifting* untuk augmentasi data, MFCC untuk ekstraksi ciri pada data audio, statistik dasar untuk mengambil *mean, median, min, max, standard deviation* pada tiap *coefficient, Min max scaler* untuk proses normalisasi dan KNN untuk metode klasifikasi.
Hasil: Karena pengujian dilakukan terpisah terhadap tiap gender maka terdapat dua model klasifikasi. Pada model pria akurasi tertinggi yang berhasil diperoleh sebesar 88,8% dan termasuk kedalam model yang *good fit.* Pada model wanita akurasi tertinggi yang berhasil diperoleh sebesar 92,5% namun model tidak mampu menglasifikasikan emosi pada data baru dengan benar. Kondisi ini disebut dengan *overfitting.* Setelah dilakukan pengujian, penyebab kondisi tersebut karena proses augmentasi *pitch shifting* sebesar satu nada pada wanita tidak mampu menyelesaikan masalah ukuran data *training* terlalu kecil dan tidak mengandung sampel data yang cukup untuk secara akurat merepresentasikan semua kemungkinan nilai data *input*.
Keaslian/ *state of the art*: Data penelitian yang digunakan pada penelitian ini tidak pernah digunakan pada penelitian sebelumnya karena data penelitian didapat dengan mengunduh dari *Youtube* kemudian diolah hingga data siap digunakan untuk penelitian.

1. Pendahuluan

Manusia dan emosi adalah satu kesatuan yang tidak bisa terpisahkan. Banyak manusia mengambil keputusan berdasarkan emosi yang sedang dirasakan. Pada umumnya untuk mengetahui kondisi emosi seseorang dapat dilihat dari raut wajahnya namun dengan perkembangan teknologi yang begitu pesat, kini emosi seseorang dapat diketahui dengan mendengar suara yang diucapkan. Kemampuan AI untuk mendeteksi emosi berdasarkan sinyal suara merupakan bagian dari *Speech Emotion Recognition* (SER), yaitu sebuah bidang studi yang bertujuan untuk mengidentifikasi keadaan emosional pembicara dari sinyal suara yang diucapkan dengan menerapkan machine learning dan pemrosesan sinyal [1].

Mel-Frequency Cepstral Coefficients (MFCC) merupakan metode ekstraksi ciri yang umum digunakan untuk pengenalan emosi pada suara karena memiliki cara kerja yang mirip dengan telinga manusia [3]. Hal ini menjadikan MFCC mampu mengklasfikasikan sebagian besar penelitian yang berhubungan dengan pengenalan berdasarkan sinyal suara [4]. Sedangkan K-Nearest Neighbor (KNN) merupakan algoritma supervised machine learning yang banyak

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v20i1.9518

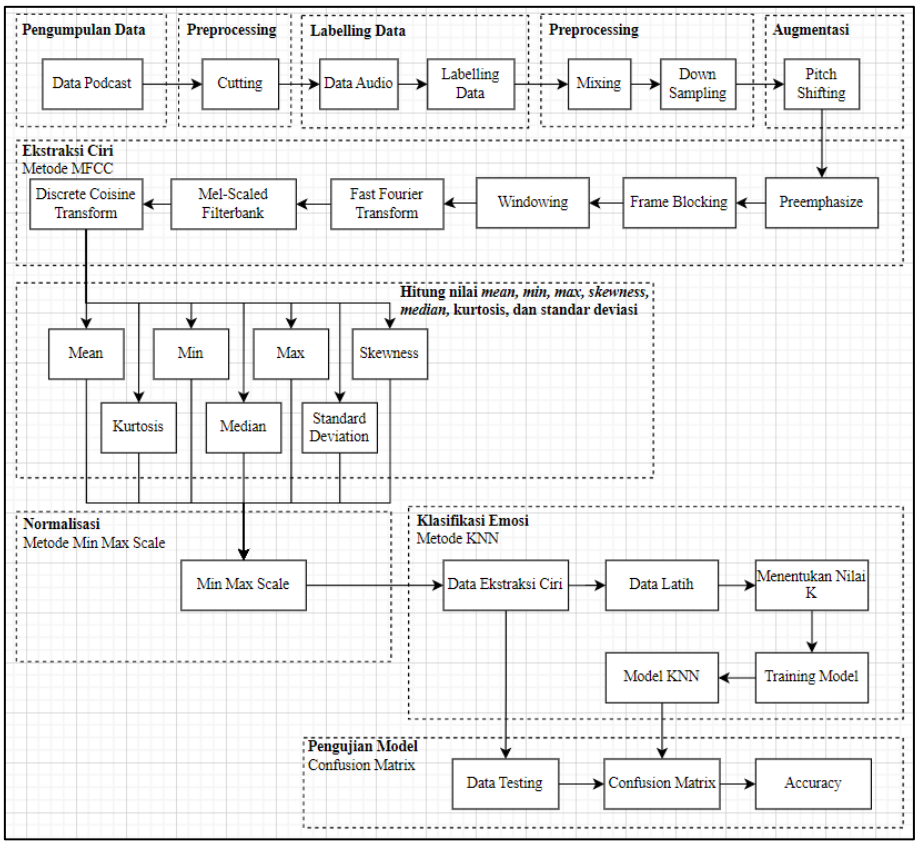
digunakan untuk menyelesaikan permasalahan klasifikasi maupun deteksi [6]. KNN memiliki kelebihan yaitu dapat diterapkan pada data dengan jumlah yang kecil dan tangguh terhadap suatu data latih yang *noise* [7].

Beberapa penelitian yang berkaitan dengan deteksi emosi berdasarkan sinyal suara pernah dilakukan oleh Hosseini menggunakan MFCC dan Gaussian Mixture Model (GMM) untuk deteksi emosi pada dataset EMO-DB (bahasa Jerman) [8]. Helmiyah menggunakan metode ekstraksi ciri MFCC dan Artificial Neural Network (ANN) untuk deteksi emosi pada dataset EMO-DB (bahasa Jerman) [3]. Al Dujaili menggunakan metode ekstraksi ciri *fundamental frequency* dan analisis spectrum, serta metode Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN) untuk deteksi emosi [2]. Al Dujaili menerapkan metode tersebut untuk mendeteksi emosi pada dataset EMO-DB (bahasa Jerman) dan SAVEE (bahasa Inggris). Dari beberapa penelitian yang telah dilakukan, mayoritas menggunakan dataset penelitian yang berbahasa asing. Sehingga penelitian ini akan menerapkan metode Mel-Frequency Cepstral Coefficients (MFCC) sebagai metode ekstraksi ciri dan metode K Nearest Neighbor (KNN) sebagai metode deteksi emosi. Data yang akan digunakan pada penelitian ini adalah data audio *podcast* berbahasa Indonesia yang diunduh dari *Youtube*. Penelitian ini akan berfokus terhadap intonasi suara, bukan berdasarkan konteks yang diucapkan. Emosi yang dipilih yaitu marah, bahagia dan sedih.

2. Metode Penelitian

Metodologi penelitian merupakan gambaran terkait alur kerja atau tahapan dalam penelitian. Alur penelitian yang dilakukan ditunjukkan pada **Gambar 1**. Tahapan dalam penelitian ini dimulai dengan proses pengambilan data yang berasal dari situs *Youtube* dengan mengunduh beberapa audio *podcast* berbahasa Indonesia. Setelah dilakukan pengambilan data kemudian data tersebut dilakukan *preprocessing* seperti *mixing, down sampling*, dan *cutting*. Data yang sudah melewati *preprocessing* selanjutnya data akan diaugmentasi untuk diperbanyak dengan proses *pitch shifting.* Tahap selanjutnya yaitu ekstraksi ciri dengan tujuan untuk mendapat ciri pada data audio dengan menggunakan metode MFCC. Hasil dari ekstraksi ciri tersebut kemudian dihitung nilai *mean, min, max, median*, dan standar deviasi pada tiap *coefficient*. Selanjutnya yaitu proses normalisasi dengan menggunakan metode *min max scale*. Proses terakhir yaitu proses klasifikasi dengan menggunakan metode KNN. Model diuji dengan menggunakan *confusion matrix* untuk mengetahui akurasi model.

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v20i1.9518



Gambar 1. Alur tahapan penelitian

2.1. Pengumpulan Data

Dataset yang digunakan pada penelitian ini merupakan data audio *podcast* berbahasa Indonesia yang diunduh dari situs *https://youtube.com*. *Youtube* dipilih karena *Youtube* merupakan sebuah *platform* yang menyediakan berbagai macam video, salah satunya video *podcast* berbahasa Indonesia. Selanjutnya video *podcast* tersebut diunduh dalam format audio. Beberapa *channel* pada situs *Youtube* yang menyediakan video *podcast* berbahasa Indonesia dan digunakan pada penelitian ini yaitu Deddy Corbuzier, CURHAT BANG Denny Sumargo, UUS Kamukita, Pita Kuning, Volix Media, Langit Entertainment, Daniel Mananta Network, dan Karin Novilda. Data audio *podcast* kemudian dipotong (*cutting*) pada bagian yang dirasa terdapat emosi (bahagia, marah atau sedih) sehingga menghasilkan data audio dengan durasi maksimal 10 detik. Data audio yang telah dipotong selanjutnya diberi label emosi oleh sarjana psikologi. Untuk data audio yang telah dipotong namun tidak termasuk kedalam emosi bahagia, marah atau sedih maka akan diberi label netral dan tidak akan digunakan dalam penelitian.

Data audio yang digunakan pada penelitian ini memiliki format *wav*, *sampling rate* 22050 Hz, dan hanya memiliki satu channel atau *mono*. Data audio diucapkan oleh pria dan wanita dewasa. Jumlah data yang digunakan pada penelitian ini berjumlah 90 data pada tiap gender sebelum proses augmentasi.

2.2. Preprocessing

Preprocessing data adalah proses pengolahan data sebelum data digunakan sebagai data latih suatu model. *Preprocessing* umumnya digunakan untuk menghilangkan *noise* pada data dan menyeragamkan data. *Preprocessing* data yang digunakan pada peneltian ini yaitu *cutting*,

mixing, dan *down sampling*. *Preprocessing* data dilakukan dengan menggunakan aplikasi *Audacity*. *Audacity* merupakan aplikasi yang memungkinkan pengguna untuk menyunting *file* audio.

2.3. Augmentasi

Augmentasi data adalah sebuah teknik atau metode yang digunakan untuk meningkatkan jumlah data dengan menambahkan salinan yang dimodifikasi dari data yang sudah ada [18]. Augmentasi yang dilakukan yaitu dengan melakukan *pitch shifting* pada data audio. Audio digeser sebesar setengah nada kearah positif dan kearah negatif. Augmentasi yang dilakukan pada penelitian ini menggunakan bantuan *library librosa*.

2.4. Mel-Frequency Cepstral Coefficient

Mel-Frequency Cepstral Coefficients (MFCC) menurut Helmiyah adalah metode ekstraksi ciri pada sinyal audio yang sering digunakan untuk pengenalan emosi pada suara karena memiliki cara kerja yang mirip dengan telinga manusia [10]. MFCC memiliki resolusi frekuensi yang baik pada frekuensi rendah (< 1000 Hz), dimana ciri sinyal ucapan lebih dominan di frekuensi tersebut [5]. MFCC terdiri dari beberapa tahap yaitu *Pre-emphasis*, *Frame-Blocking*, *Windowing*, *Discrete Fourier Transform*, *Mel-scaled Filterbank*, dan *Discrete Cosine Transform*.

1. *Pre-emphasis*

   Pre-emphasis adalah proses yang dilakukan untuk meningkatkan kualitas sinyal dan meningkatkan akurasi pada saat ekstraksi ciri dengan cara penekanan sinyal pada frekuensi tinggi sehingga selaras dengan frekuensi rendah [9].

   $y(n) = s(n) - \alpha\, s(n-1)$ ..... (2.1)

   Keterangan:
   y(n) : sinyal hasil pre-emphasis pada data ke-n
   s(n) : nilai sinyal pada data ke-n
   α : konstanta filter pre-emphasis

2. *Frame Blocking*

   Frame blocking adalah proses dimana sinyal dibagi menjadi beberapa *frame* kecil. *Frame blocking* penting untuk dilakukan karena sinyal harus diproses dalam satuan waktu tertentu (*short time*), karena sinyal suara terus berubah seiring waktu [11].

   $\#F = \left|\frac{N_s - H_l}{N_f - H_l}\right|$ ..... (2.2)

   Keterangan:
   $\#F$: jumlah *frame*
   N_S : jumlah data dalam suatu sinyal

3. *Windowing*

   Windowing adalah sebuah proses mengalikan fungsi *windowing* dengan sampel pada tiap *frame* dengan tujuan untuk menutupi kekurangan ketika proses mengubah data dari domain waktu menjadi domain frekuensi yang biasa disebut dengan batas diskontinuitas (*boundary of discontinues*) [12]. Persamaan *windowing* secara umum dapat dilihat pada persamaan 2.3.

   $y_w(n) = x(n) \times w(n)$ ..... (2.3)

Keterangan:
$y_w(n)$: sinyal hasil proses *windowing* ke-n
$x(n)$: nilai sinyal pada data ke-n
$w(n)$: nilai fungsi *window* pada data ke-n

4. *Fast Fourier Transform*

Fast Fourier Transform (FFT) adalah pengembangan dari algoritma *Discrete Fourier Transform* (DFT) yang memiliki fungsi untuk mengonversi sinyal dari domain waktu menjadi domain frekuensi [12]. DFT memiliki kekurangan yaitu membutuhkan waktu komputasi yang terlalu lama dan tidak efisien sehingga muncul metode FFT untuk menutupi kekurangan tersebut [13]. *Short Time Fourier Transform* (STFT) merupakan salah satu metode FFT yang banyak digunakan.

$$S(m,k) = \sum_{n=0}^{N_f-1} x_w(n,m) \times e^{-i2\pi n\frac{k}{N}}; k = 1,2,3,\dots,\frac{N_f}{2}+1 \quad \text{..... (2.4)}$$

Keterangan:
$S(m,k)$: hasil perhitungan *fourier transform* ke-k pada *frame* ke-m
$x_w(n,m)$: sinyal hasil *windowing* ke-n pada *frame* ke-m
N_f : jumlah data dalam satu *frame*
k : variabel *frequency discrete* (*frequency bin*)
n : data sinyal ke-n

5. *Mel-Scaled Filterbank*

Mel-scaled filterbank adalah proses *filtering* spektrum pada tiap *frame*. Filter ini digunakan untuk mengikuti persepsi telinga manusia dalam menerima frekuensi suara [14]. Pemetaan antara frekuensi (Hz) dengan skala mel linier ketika di bawah 1000 Hz dan logaritmik ketika di atas 1000 Hz.

$$mel(f) = 2595\ log_{10}(1+\frac{f}{700}) \quad \text{..... (2.5)}$$

Keterangan:
f : *frequency*

6. *Discrete Cosine Transform*

Discrete Coisine Transform (DCT) adalah langkah terakhir dari proses utama ekstraksi ciri MFCC. DCT pada dasarnya memiliki kesamaan dengan *inverse fourier transform*. Namun, hasil yang didapat dari proses DCT lebih mendekati *Principal Component Analysis* (PCA). PCA adalah metode statistik klasik yang digunakan secara luas dalam analisis data dan kompresi [13].

$$C_m = \sum_{k=1}^{K} log_{10}\, Y[k] cos\left[m(k-\frac{1}{2})\frac{\pi}{K}\right]; m = 1,2,\dots,K \quad \text{..... (2.6)}$$

Keterangan:

C_m : *coefficient* MFCC
$Y[k]$: hasil dari proses *filterbank* pada indeks ke-*k*
m : nilai *coefficient*
K : jumlah *coefficient* yang diinginkan

2.5. Statistik Dasar

Setelah melakukan ekstraksi ciri pada sinyal audio, data yang semula berupa sinyal audio berganti menjadi data hasil ekstraksi ciri MFCC. Banyaknya data yang dihasilkan dari proses ekstraksi ciri bergantung pada jumlah *coefficient* MFCC dan durasi audio suara ucapan. Data pada tiap *coefficient* selanjutnya dihitung untuk mendapat nilai statistik, nilai statistik ini yang kemudian digunakan untuk melatih model. Statistik dasar yang digunakan pada penelitian ini yaitu, *mean, max, median, min,* dan standar deviasi.

2.6. Normalisasi

Normalisasi data merupakan sebuah proses untuk menentukan nilai supaya nilai tersebut berada pada range tertentu. Selain itu normalisasi juga memiliki tujuan untuk mempertahankan kualitas model agar tetap baik terhadap standar deviasi fitur atau ciri yang sangat kecil [15]. Normalisai dilakukan dengan menggunakan *min max scale.*

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (2.7)$$

Keterangan:

x' : Nilai data setelah normalisasi
x : Nilai data sebelum normalisasi
$\min(x)$: Nilai minimum dari kumpulan data x
$\max(x)$: Nilai maksimum dari kumpulan data x

2.7. K-Nearest Neighbor

K Nearest Neighbor (KNN) adalah metode klasifikasi yang bekerja dengan cara mencari jarak terdekat antara data baru yang belum diketahui class-nya dengan k tetangga (*neighbor*) terdekatnya berdasarkan data latih. Jarak antara data baru dengan data latih dihitung dengan cara mengukur jarak antara titik yang merepresentasikan data baru dengan semua titik yang merepresentasikan data latih dengan rumus *Euclidean Distance* [16]. Nilai k umumnya ditentukan dalam bilangan ganjil untuk menghindari terjadinya jumlah jarak yang sama dalam proses klasifikasi [17].

Algoritma KNN pada dasarnya diimplementasikan dalam langkah-langkah berikut:

1. Tentukan jumlah *k* yaitu jumlah tetangga terdekat
2. Hitung jarak data baru dengan semua data latih
3. Urutkan jarak tersebut dan tetapkan tetangga terdekat berdasarkan jarak minimum ke-*k*
4. Periksa *class* tetangga terdekat
5. Data baru akan diklasifikasikan berdasarkan mayoritas *class* tetangga terdekat

$$d = \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2} \quad (2.8)$$

Keterangan:

d : *Euclidean distance*
a_i : data uji pada indeks ke-i
b_i : data latih pada indeks ke-i

3. Hasil dan Pembahasan

3.1. Hasil Pengujian Model

Pengujian metode bertujuan untuk mengetahui tingkat akurasi metode dalam mengklasifikasikan emosi berdasarkan suara ucapan. Pengujian antara gender laki-laki dan

perempuan dilakukan secara terpisah karena laki-laki dan perempuan memiliki karateristik suara yang berbeda.

Akurasi terbaik yang berhasil diraih oleh model untuk gender pria dari proses parameter *tuning* sebesar 88,8% untuk $k = 3$. Akurasi ini masih diperlu diuji untuk mengetahui seberapa mampu model dalam mengklasifikasikan emosi dari data baru. Data baru yang dimaksud pada penelitian ini yaitu data ucapan (*podcast*) yang diucapkan oleh narasumber baru. Data yang akan digunakan untuk pengujian ini berjumlah 30 data, dengan detail 10 data tiap emosi. Tabel pengujian model pada pria dapat dilihat pada **Tabel 1**. Dari 30 data yang diuji, model mampu mengklasifikasikan emosi suara ucapan dengan benar sebanyak 24 data. Emosi marah yang berhasil diklasifikasikan berjumlah 7 dari 10 data. Emosi bahagia yang berhasil diklasifikasikan berjumlah 8 dari 10 data. Emosi sedih yang berhasil diklasfikasikan berjumlah 9 dari 10 data.

Tabel 1. Hasil Pengujian Model pada Pria

No.	Nama *file*	Emosi yang diharapkan	Emosi yang didapatkan
1.	IG 3	Marah	Bahagia
2.	IG 4	Marah	Bahagia
3.	IG 5	Marah	Marah
4.	RA 6	Marah	Marah
5.	RA 11	Marah	Marah
6.	S 4	Marah	Bahagia
7.	S 5	Marah	Marah
8.	M 1	Marah	Marah
9.	J 1	Marah	Marah
10.	J 2	Marah	Marah
11.	PT 1	Bahagia	Bahagia
12.	PT 3	Bahagia	Bahagia
13.	IG 6	Bahagia	Marah
14.	IG 7	Bahagia	Bahagia
15.	RA 1	Bahagia	Marah
16.	RA 2	Bahagia	Bahagia
17.	S 1	Bahagia	Bahagia
18.	S 2	Bahagia	Bahagia
19.	CP 1	Bahagia	Bahagia
20.	CP 2	Bahagia	Bahagia
21.	RH 1	Sedih	Sedih
22.	RH 2	Sedih	Sedih
23.	VJ 1	Sedih	Sedih
24.	VJ 2	Sedih	Sedih
25.	FS 1	Sedih	Sedih
26.	FS 2	Sedih	Sedih
27.	DM 1	Sedih	Sedih
28.	DM 2	Sedih	Sedih
29.	AP 2	Sedih	Bahagia
30.	AP 3	Sedih	Sedih

Akurasi terbaik yang berhasil diraih oleh model untuk gender wanita dari proses parameter tuning pada wanita sebesar 92,5%. Hasil pengujian model pada wanita dapat dilihat pada 2. Dari 30 data yang diuji, model mampu mengklasifikasikan emosi suara ucapan dengan benar sebanyak 15 data. Emosi marah yang berhasil diklasifikasikan berjumlah 4 dari 10 data. Emosi bahagia yang berhasil diklasifikasikan berjumlah 2 dari 10 data. Emosi sedih yang berhasil

diklasfikasikan berjumlah 9 dari 10 data. Dari keseluruhan data baru yang diuji, model hanya mampu menglasifikasikan emosi sedih.

Tabel 2. Hasil Pengujian Model pada Wanita

No.	Nama *file*	Emosi yang diharapkan	Emosi yang didapatkan
1.	MA 1	Marah	Sedih
2.	MA 3	Marah	Sedih
3.	MA 4	Marah	Sedih
4.	MA 6	Marah	Sedih
5.	NM 1	Marah	Marah
6.	NS 1	Marah	Bahagia
7.	NS 2	Marah	Marah
8.	NS 3	Marah	Marah
9.	NS 4	Marah	Marah
10.	F 2	Marah	Bahagia
11.	AT 1	Bahagia	Marah
12.	AT 2	Bahagia	Marah
13.	MA 2	Bahagia	Sedih
14.	PL 1	Bahagia	Marah
15.	PL 2	Bahagia	Sedih
16.	PL 3	Bahagia	Bahagia
17.	PL 4	Bahagia	Bahagia
18.	RS 1	Bahagia	Marah
19.	RS 2	Bahagia	Marah
20.	RS 3	Bahagia	Marah
21.	MA 5	Sedih	Sedih
22.	MA 7	Sedih	Marah
23.	CT 1	Sedih	Sedih
24.	CT 2	Sedih	Sedih
25.	CT 3	Sedih	Sedih
26.	GE 1	Sedih	Sedih
27.	GE 2	Sedih	Sedih
28.	GE 3	Sedih	Sedih
29.	JI 1	Sedih	Sedih
30.	JI 2	Sedih	Sedih

3.2. Pembahasan

Terdapat dua pengujian yang dilakukan pada penelitian ini, pengujian pertama adalah *parameter tuning* dan pengujian kedua adalah pengujian kemampuan model. Pengujian dilakukan untuk masing-masing gender. Pengujian kedua yaitu menguji kemampuan model. Model *machine learning* yang sebelumnya memiliki akurasi tertinggi diuji dengan data baru untuk mengetahui apakah model sudah cukup baik dalam menglasifikasikan emosi berdasarkan suara ucapan. Data baru yang dimaksud dalam pengujian ini adalah data suara ucapan narasumber yang tidak digunakan kedalam data latih. Data baru diambil dari beberapa video *podcast* di *youtube*. Karena pengujian dilakukan secara terpisah pada tiap gender, maka yang pertama akan dibahas adalah pengujian model pada pria.

Dari proses *parameter tuning* yang telah dilakukan, akurasi tertinggi model pada pria yang berhasil diperoleh sebesar 88,8%. Selanjutnya model diuji dengan data baru, dari 30 data baru yang diuji, model mampu mengklasifikasikan emosi suara ucapan dengan benar sebanyak 24 data. Emosi marah yang berhasil diklasifikasikan berjumlah 7 dari 10 data. Emosi bahagia yang

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v20i1.9518

berhasil diklasifikasikan berjumlah 8 dari 10 data. Emosi sedih yang berhasil diklasfikasikan berjumlah 9 dari 10 data. Dengan hasil pengujian yang mendekati akurasi awal maka model yang dibuat tergolong kedalam *good fit.*

Pengujian model selanjutnya adalah pengujian model pada wanita. Dari proses *parameter tuning* yang telah dilakukan, akurasi tertinggi model pada wanita yang berhasil diperoleh sebesar 92,5%. Dari 30 data baru yang diuji, model mampu mengklasifikasikan emosi suara ucapan dengan benar sebanyak 15 data. Emosi marah yang berhasil diklasifikasikan berjumlah 4 dari 10 data. Emosi bahagia yang berhasil diklasifikasikan berjumlah 2 dari 10 data. Emosi sedih yang berhasil diklasfikasikan berjumlah 9 dari 10 data. Dengan selisih hasil pengujian model dan akurasi *training* maka model yang dibuat tergolong kedalam *overfit.* Hal ini dapat terjadi ketika model mampu dengan baik menglasifikasikan data tes namun tidak mampu menglasifikasikan data baru dengan benar.

Overfit yang terjadi pada model wanita disebabkan karena data hasil augmentasi *pitch shifting* tidak mengandung sampel data yang cukup untuk secara akurat merepresentasikan semua kemungkinan nilai data *input.* Dimana *pitch shifting* yang dilakukan hanya menggeser nada sebesar satu nada.

4. Kesimpulan dan Saran

Berdasarkan penelitian yang telah dilakukan, didapatkan kesimpulan yang dapat dinyatakan sebagai berikut. Kombinasi antara metode ekstraksi ciri MFCC dan metode KNN mampu dalam mendeteksi emosi berdasarkan intonasi suara pada pria dengan akurasi sebesar 88,8% terhadap data *test* dan 80% terhadap data baru. MFCC dan KNN juga mampu mendeteksi emosi berdasarkan intonasi suara pada wanita dengan akurasi 92,5% terhadap data *test* namun terhadap data baru hanya mampu mencapai akurasi sebesar 50%. Model pada wanita mengalami kondisi *overfitting* yang disebabkan oleh data augmentasi *pitch shifting* yang tidak mampu menghasilkan sampel data yang cukup untuk secara akurat merepresentasikan semua kemungkinan nilai data *input.*

Adapun saran yang dapat diberikan untuk penelitian selanjutnya yaitu:

a. Jumlah data yang digunakan ditambah sehingga lebih beragam dan model lebih baik dalam menglasifikasikan emosi.
b. Emosi yang digunakan pada penelitian ini merupakan emosi dasar, namun pada kenyataannya emosi manusia beragam. Sehingga pada penelitian berikutnya dapat menggunakan data dengan emosi yang lebih beragam.
c. Menggunakan kombinasi metode ekstrasi ciri dan metode klasifikasi yang lain untuk menghasilkan model yang lebih baik.

Daftar Pustaka

[1] Alghifari, M. F., Gunawan, T. S., & Kartiwi, M. (2018). Speech Emotion Recognition Using Deep Feedforward Neural Network. Indonesian Journal of Electrical Engineering and Computer Science, 10(2), 554–561. https://doi.org/10.11591/ijeecs.v10.i2.pp554-561

[2] Al Dujaili, M. J., Ebrahimi-Moghadam, A., & Fatlawi, A. (2021). Speech emotion

recognition based on SVM and KNN classifications fusion. International Journal of Electrical and Computer Engineering, 11(2), 1259–1264. https://doi.org/10.11591/ijece.v11i2.pp1259-1264

[3] Helmiyah, S., Riadi, I., Umar, R., & Hanif, A. (2021). Speech Classification to Recognize Emotion Using Artificial Neural Network. Khazanah Informatika: Jurnal Ilmu Komputer Dan Informatika, 7(1), 12–17. https://doi.org/10.23917/khif.v7i1.11913

[4] Liu, G., He, W., & Jin, B. (2018). Feature Fusion of Speech Emotion Recognition Based on Deep Learning. Proceedings of IC-NIDC.

[5] Aini, Y. K., Santoso, T. B., & Dutono, D. T. (2021). Pemodelan CNN Untuk Deteksi Emosi Berbasis Speech Bahasa Indonesia. Jurnal Komputer Terapan, 7(1), 143–152. https://jurnal.pcr.ac.id/index.php/jkt/

[6] Albon, C., 2018. Machine Learning with Python Cookbook. Sebastopol: O'Reilly Media.

[7] Mahardika, Kukuh W., Sari, Yuita A., & Arwan, Achmad. (2018). Optimasi K-Nearest Neighbour Menggunakan Particle Swarm Optimization Optimasi K-Nearest Neighbour Menggunakan Particle Swarm Optimization pada Sistem Pakar untuk Monitoring Pengendalian Hama pada Tanaman Jeruk. Jurnal Teknologi, 2(July), 13.

[8] Hosseini, Z., Ahadi, S. M., & Faraji, N. (2014). Speech Emotion Classification via a Modified Gaussian Mixture Model Approach. 2014 7th International Symposium on Telecommunications, IST 2014, 487–491. https://doi.org/10.1109/ISTEL.2014.7000752

[9] Arifin, C., & Junaedi, H. (2018). Emotion Sound Classification with Support Vector Machine Algorithm. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 3(2), 181–190. https://doi.org/10.22219/kinetik.v3i2.610

[10] Helmiyah, S., Riadi, I., Umar, R., Hanif, A., Yudhana, A., & Fadlil, A. (2020). Identifikasi Emosi Manusia Berdasarkan Ucapan Menggunakan Metode Ekstraksi Ciri LPC dan Metode Euclidean Distance. Jurnal Teknologi Informasi Dan Ilmu Komputer, 7(6), 1177. https://doi.org/10.25126/jtiik.2020722693

[11] Putra, K. T. (2017). Sistem Pengenal Wicara Menggunakan Mel-Frequency Cepstral Coefficient (Speech Recognition System Using Mel-Frequency Cepstral Coefficient). Semesta Teknika, 20(1), 75–80.

[12] Krishna Kishore, K. V., & Krishna Satish, P. (2013). Emotion recognition in speech using MFCC and wavelet features. Proceedings of the 2013 3rd IEEE International Advance Computing Conference, IACC 2013, 842–847. https://doi.org/10.1109/IAdCC.2013.6514336

[13] Heriyanto, H., Hartati, S., & Putra, A. E. (2018). Ekstraksi Ciri Mel Frequency Cepstral Coefficient (Mfcc) Dan Rerata Coefficient Untuk Pengecekan Bacaan Al-Qur'an. Telematika, 15(2), 99. https://doi.org/10.31315/telematika.v15i2.3123

[14] Muljono, Prasetya, M. R., Harjoko, A., & Supriyanto, C. (2019). Speech Emotion Recognition of Indonesian Movie Audio Tracks based on MFCC and SVM. Proceedings of the 4th International Conference on Contemporary Computing and Informatics, IC3I

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v20i1.9518

2019, 22–25. https://doi.org/10.1109/IC3I46837.2019.9055509

[15] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

[16] Hermawan, Y. D., Hariadi, V., & Amaliah, B. (2017). Implementasi Algoritma K-Nearest Neighbors dengan Particle Swarm Optimization dalam Klasifikasi Trouble pada Base Transceiver Station (BTS). Jurnal Teknik ITS.

[17] Harsemadi, G., Sudarma, M., & Pramaita, N. (2017). Implementasi Algoritma K-Nearest Neighbor pada Perangkat Lunak Pengelompokan Musik untuk Menentukan Suasana Hati. Majalah Ilmiah Teknologi Elektro, 16(1), 14–20. https://doi.org/10.24843/mite.1601.03

[18] Nurcahyo, R., & Iqbal, M. (2022). Pengenalan Emosi Pembicara Menggunakan Convolutional Neural Networks. Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), 6(1), 115– 122. https://doi.org/10.29207/resti.v6i1.3726