Jurnal Inovasi Pembelajaran Progresif

https://ejurnals.com/ojs/index.php/jipp Vol. 6, No. 3, September 2025

DETEKSI EMOSI BERBASIS SUARA MENGGUNAKAN K-NEAREST NEIGHBORS DAN EKSTRAKSI MFCC

Mendri Syaputra[1], Muhamad Akbar[2], Taqwa Martadinata[3]

[1,2,3]Universitas Bina Insan

Email: Mendrisyaputra01@gmail.com[1], muhamad.akbar@univbinainsan.ac.id[2], taqwa@univbinainsan.ac.id[3]

Abstrak: Penelitian ini bertujuan mengembangkan sistem deteksi emosi berbasis suara dengan metode K-Nearest Neighbors (KNN) dan ekstraksi fitur Mel Frequency Cepstral Coefficient (MFCC). Dataset terdiri dari rekaman suara yang merepresentasikan beberapa jenis emosi. Fitur suara diekstraksi menggunakan MFCC untuk mendapatkan representasi spektral yang lebih kompak. KNN digunakan sebagai algoritma klasifikasi, dan pengujian dilakukan dengan variasi nilai k. Hasil terbaik diperoleh pada k=9 dengan akurasi 90%, menunjukkan kombinasi metode ini efektif untuk mendeteksi emosi suara.
Kata Kunci: Deteksi Emosi, KNN, MFCC, Pengolahan Sinyal Suara.

Abstract: *This research aims to develop a voice-based emotion detection system using the K-Nearest Neighbors (KNN) method and Mel Frequency Cepstral Coefficient (MFCC) feature extraction. The dataset contains audio recordings representing various emotion classes. MFCC is employed to extract compact spectral features from the audio signals. KNN is applied for classification, and experiments were conducted with different k values. The best performance was achieved at k=9 with an accuracy of 90%, indicating that this method combination is effective for speech emotion recognition.*
Keywords: *Emotion Detection, KNN, MFCC, Speech Signal Processing.*

PENDAHULUAN

Kemampuan sistem untuk memahami emosi manusia melalui suara telah menjadi bidang penelitian yang semakin penting dalam ranah pengolahan sinyal digital dan kecerdasan buatan. Emosi merupakan aspek fundamental dalam kehidupan manusia yang berperan besar dalam pengambilan keputusan, pembentukan perilaku, dan interaksi sosial. Dalam komunikasi sehari-hari, emosi bukan hanya sarana untuk mengekspresikan perasaan, tetapi juga menjadi penentu dalam penyampaian maksud, intonasi, dan konteks pembicaraan. Hal ini menjadikan penelitian mengenai deteksi emosi berbasis suara (Speech Emotion Recognition/SER) relevan untuk terus dikembangkan, terlebih di era modern yang menuntut sistem interaktif semakin cerdas, responsif, dan adaptif terhadap kondisi emosional penggunanya.

Kemajuan teknologi multimedia dan kecerdasan buatan kini memungkinkan komputer untuk mengenali dan merespons emosi pengguna secara otomatis. Implementasi sistem deteksi emosi berbasis suara dapat diterapkan pada berbagai bidang, misalnya pada asisten virtual yang dapat menyesuaikan respons berdasarkan suasana hati pengguna, layanan pelanggan yang mampu mendeteksi tingkat kepuasan atau kekecewaan konsumen, serta sistem pembelajaran adaptif yang mampu menyesuaikan materi berdasarkan kondisi emosional pelajar. Bahkan dalam bidang kesehatan, deteksi emosi suara dapat digunakan untuk membantu terapi psikologis atau pemantauan kondisi mental pasien. Dengan semakin luasnya potensi aplikasi, penelitian ini diharapkan dapat berkontribusi pada pengembangan teknologi yang mendukung interaksi manusia–mesin secara lebih manusiawi.

Berbagai pendekatan telah digunakan dalam penelitian terdahulu untuk mendeteksi emosi suara. Metode yang populer meliputi Support Vector Machine (SVM), Deep Neural Network (DNN), Convolutional Neural Network (CNN), serta metode tradisional seperti K-Nearest Neighbors (KNN). Pendekatan berbasis deep learning memang menawarkan tingkat akurasi yang tinggi, namun membutuhkan sumber daya komputasi yang besar dan waktu pelatihan yang lama. Sebaliknya, KNN menawarkan kesederhanaan, transparansi, serta kebutuhan komputasi yang lebih rendah, sehingga lebih praktis untuk penelitian berskala kecil hingga menengah. KNN bekerja dengan prinsip jarak terdekat antara data uji dengan data latih, sehingga klasifikasi dilakukan berdasarkan mayoritas tetangga terdekat. Keunggulan ini membuat KNN masih relevan digunakan, terutama untuk penelitian awal yang menekankan pada kecepatan pengujian dan kemudahan implementasi.

Selain algoritma klasifikasi, teknik ekstraksi fitur juga memiliki peran yang sangat penting. Ekstraksi fitur bertujuan untuk menyederhanakan sinyal suara menjadi representasi numerik yang tetap mempertahankan ciri khas emosi. Salah satu metode ekstraksi fitur yang banyak digunakan adalah Mel Frequency Cepstral Coefficient (MFCC). MFCC didasarkan pada skala Mel, yaitu skala frekuensi yang dirancang untuk meniru cara telinga manusia merespons perbedaan frekuensi [2]. Dengan demikian, MFCC mampu menangkap karakteristik suara seperti pitch, formant, energi, serta durasi pengucapan yang sangat berkaitan dengan ekspresi emosi. Proses MFCC melibatkan beberapa tahap, seperti pre-emphasis filtering, framing, windowing, Fast Fourier Transform (FFT), hingga transformasi

kosinus diskrit (DCT) [3]. Hasil akhir berupa koefisien yang mencerminkan informasi spektral suara, dan inilah yang kemudian digunakan sebagai input dalam proses klasifikasi

Dalam konteks penelitian bahasa Indonesia, studi tentang Speech Emotion Recognition masih relatif terbatas. Salah satu kendala utamanya adalah minimnya ketersediaan dataset standar yang memuat variasi suara dengan label emosi yang jelas. Kebanyakan penelitian masih menggunakan dataset berbahasa Inggris atau bahasa asing lainnya, seperti Toronto Emotional Speech Set (TESS), karena lebih mudah diakses dan sudah terstandarisasi.

Akan tetapi, penggunaan dataset asing dapat menjadi keterbatasan tersendiri karena setiap bahasa memiliki ciri fonetis dan intonasi yang berbeda. Tantangan lainnya adalah bagaimana sistem dapat memahami hubungan kompleks antara gaya berbicara dengan emosi. Seseorang yang sedang marah, misalnya, dapat mengekspresikan emosi dengan suara keras, intonasi tinggi, dan tempo cepat, sedangkan emosi sedih sering ditandai dengan nada rendah, intensitas lemah, serta tempo lambat. Variasi inilah yang perlu dipahami oleh sistem agar mampu melakukan klasifikasi secara akurat.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengimplementasikan sistem deteksi emosi berbasis suara menggunakan algoritma KNN dengan ekstraksi fitur MFCC. Fokus utama penelitian adalah mengoptimalkan parameter k pada algoritma KNN agar diperoleh akurasi tertinggi. Nilai k sangat menentukan hasil klasifikasi, sehingga perlu diuji beberapa konfigurasi untuk mendapatkan nilai terbaik. Dataset yang digunakan terdiri dari sejumlah data suara yang telah terklasifikasi ke dalam kategori emosi tertentu, seperti senang, marah, sedih, takut, dan jijik. Dengan kombinasi MFCC dan KNN, penelitian ini diharapkan dapat menghasilkan model yang mampu mengenali emosi manusia secara akurat, stabil, dan konsisten.

Kontribusi penelitian ini tidak hanya pada tataran teoritis, tetapi juga praktis. Dari sisi teoritis, penelitian ini menambah referensi dalam bidang pengolahan sinyal suara dan kecerdasan buatan, khususnya terkait pemanfaatan KNN dan MFCC dalam deteksi emosi. Dari sisi praktis, hasil penelitian dapat dijadikan dasar dalam pengembangan sistem interaktif yang lebih cerdas, misalnya untuk chatbot layanan pelanggan yang peka terhadap emosi, sistem pembelajaran daring yang adaptif terhadap kondisi emosional siswa, hingga perangkat pendukung kesehatan mental. Dengan demikian, penelitian ini diharapkan mampu membuka peluang untuk inovasi lebih lanjut di bidang teknologi interaksi manusia–mesin.

METODE PENELITIAN

1) Metode Penelitian

Metode penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan algoritma *Machine Learning* untuk melakukan klasifikasi emosi berbasis suara. Penelitian dilaksanakan melalui serangkaian tahapan mulai dari pengumpulan data, persiapan data, ekstraksi fitur, hingga tahap klasifikasi dan evaluasi model. Untuk menggambarkan alur kerja, digunakan kerangka penelitian berbasis model Cross-Industry Standard Process for Data Mining (CRISP-DM) yang meliputi enam tahapan utama, yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*.

a. Business Understanding

Tahapan ini bertujuan untuk mengembangkan sistem yang mampu mendeteksi emosi manusia berdasarkan suara menggunakan algoritma K- Nearest Neighbor (KNN) dan fitur Mel Frequency Cepstral Coefficient (MFCC).

b. Data Understanding

Pada tahapan Data Understanding, proses dimulai dengan identifikasi dan pengumpulan dataset yang relevan untuk tujuan deteksi emosi suara. Tahap Data Understanding dilakukan untuk memperoleh gambaran awal terhadap dataset suara yang digunakan.

c. Data Preparation Tahap

Data Preparation (persiapan data) bertujuan untuk mengolah dan mempersiapkan data agar siap digunakan dalam proses pelatihan dan pengujian model. Tahapan ini merupakan langkah krusial karena kualitas data yang digunakan akan sangat memengaruhi akurasi hasil klasifikasi.

d. Modeling

Tahap modeling merupakan proses pembangunan model klasifikasi berdasarkan data yang telah dipersiapkan. Tujuan utama dari tahap ini adalah untuk membuat model yang mampu memprediksi kelas emosi dari suara berdasarkan fitur yang telah diekstraksi sebelumnya.

e. Evaluation

Pada tahapan Evaluation, model yang telah dibangun di evaluasi untuk mengukur kinerja model dalam mendeteksi emosi berbasis 22 suara Evaluasi yang dipakai yaitu Confusion Matrix.

f. Deployment

Pada tahapan deployment model deteksi emosi suara dapat digunakan secara praktis. Model yang telah dilatih disimpan dalam format .pkl dan diintegrasikan ke dalam antarmuka sederhana berbasis Python, sehingga pengguna dapat memasukkan suara dan langsung melihat hasil klasifikasinya. Sistem juga dirancang agar mampu menerima input suara secara real-time dari mikrofon.

2) Metode Pengupulan Data

Pengumpulan data dalam penelitian ini dilakukan dengan Studi pustaka dan Data sekunder. Adapun penjelasan lebih terperinci adalah sebagai berikut:

a. Studi Pustaka

Studi pustaka (library research) yaitu metode dengan pengumpulan data dengan cara memahami dan mempelajari teori-teori dari berbagai literature yang berhubungan dengan penelitian tersebut.

b. Data Sekunder

Data yang digunakan dalam penelitian ini ialah data sekunder yang diperoleh secara tidak langsung melalui sumber pihak ketiga yaitu dataset yang digunakan berasal dari Toronto Emotional Speech Set dapat (TESS). Dataset TESS dikumpulkan dan diunggah melalui Kaggle, yang diakses melalui tautan berikut: https://www.kaggle.com/datasets/ejlok1/toronto-emotional-speech-set tess.

3) Metode Pengujian dan Pengolahan Data

a. Metode Pengujian

Dalam penelitian ini metode pengujian dilakukan menggunakan algoritma K-Nearest Neighbors (KNN) dalam mendeteksi emosi berbasis suara. Evaluasi ini bertujuan untuk mengetahui seberapa baik model dalam mengklasifikasikan data dengan benar. Untuk mengukur kinerja model, digunakan beberapa metrik evaluasi yang umum dalam

machine learning, yaitu:

1. Confusion Matrix

Kinerja suatu klasifikasi model dapat diukur melalui beberapa parameter pengukuran, seperti tingkat akurasi , recall, dan presisi. Untuk menghitung parameter –parameter tersebut , diperlukan sebuah matriks yang disebut Confusion Matrix.

Tabel 3.2 Confusion Matrix

Aktual	Predicted Condition	
	Positif	Negatif
Positif	True Positive (TP)	False Negative (FN)
Negatif	False Positive (FP)	True Negative (TN)

2. Akurasi (Accuracy)

Akurasi mengukur seberapa sering model berhasil memprediksi sesuatu dengan benar dibandingkan dengan jumlah total data yang diujikan [4]. Rumus untuk menghitung akurasi adalah:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

3. Presisi (Precision)

Presisi mengukur seberapa banyak prediksi positif yang benar dibandingkan dengan seluruh prediksi positif yang dihasilkan model. Rumus presisi adalah:

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

4. Recall (Sensitivitas)

Recall mengukur seberapa baik model dalam mendeteksi kelas positif dari semua sampel yang sebenarnya positif. Rumus recall adalah:

$$Recall = \frac{TP}{TP + FP} \quad (4)$$

5. F1-Score

F1-Score adalah rata-rata harmonis antara presisi dan recall, yang digunakan untuk menyeimbangkan kedua metrik tersebut. Rumus F1 Score adalah:

$$F1 - Score = 2 \text{ x } \frac{Precision \text{ x } Recall}{Precision + Recall} \quad (5)$$

b. Metode Pengolahan Data

a. Tahap Pengumpulan Data

Dataset yang digunakan terdiri dari 2.800 file audio yang telah melalui proses pembersihan dan diberi label. Data ini disusun sedemikian rupa sehingga mencakup rekaman dari dua aktor wanita berusia 26 dan 64 tahun. Setiap emosi disimpan dalam folder terpisah, yang di dalamnya berisi file audio dengan 200 kata target. Seluruh file audio tersebut menggunakan format WAV.

Gambar 3.1. Dataset

b. Tahap Preprocessing

Pada proses ini, data yang tersedia akan melalui tahap pembersihan untuk meningkatkan kualitas dan mengurangi kemungkinan error. Untuk memaksimalkan hasil, dilakukan preprocessing pada data suara dengan bantuan library Librosa dan NumPy. Beberapa teknik preprocessing yang diterapkan meliputi pengurangan noise, stretching, shifting, dan penyesuaian pitch.

1. Pitch shifting

Pitch shifting adalah teknik mengubah tinggi nada (frekuensi fundamental) dari sinyal suara tanpa mengubah durasi waktunya.

2. Time Stretching

Time stretching merupakan teknik mengubah kecepatan atau durasi audio tanpa mengubah tinggi nadanya.

3. Time Shifting

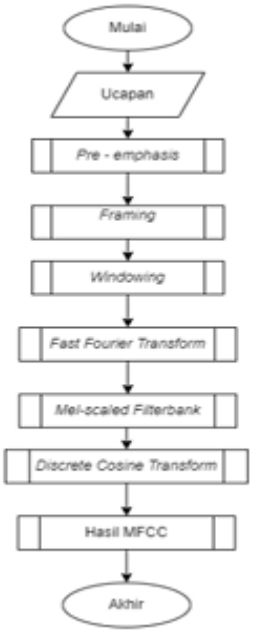
Time Shifting adalah teknik menggeser sinyal audio kedepan atau kebelakang di sepanjang sumbu waktu. Proses ini tidak mengubah isi suara, namun hanya memindahkan posisi mulai suara.

4. Pengurangan Noise

Pengurangan Noise merupakan proses menghilangkan atau mengurangi suara gangguan latar belakang (seperti dengungan, desis, atau suara lingkungan) dari sinyal suara utama.

c. Ekstraksi Fitur dengan MFCC

Setelah melalui tahapan preprocessing untuk memastikan kualitas sinyal suara yang digunakan dalam penelitian ini berada dalam kondisi optimal, langkah selanjutnya adalah melakukan ekstraksi ciri menggunakan metode Mel-Frequency Cepstral Coefficient (MFCC).



Gambar 3.2. Alur MFCC [5].

HASIL DAN PEMBAHASAN

Hasil Penelitian

a. Pengumpulan Data

Dalam penelitian ini, data utama yang digunakan berupa rekaman suara manusia yang mengandung berbagai ekspresi emosi. Sumber data terdiri dari beberapa kategori emosi dasar, yaitu takut, marah, senang, sedih, jijik, dan netral**.** Data suara ini dikumpulkan dalam format **.**wav, yang dipilih karena memiliki kualitas audio yang lebih baik dan relatif stabil dibandingkan format audio terkompresi seperti MP3. Setiap rekaman berdurasi antara 2 hingga 5 detik, sehingga cukup untuk mewakili ciri khas dari intonasi dan pola suara setiap emosi. Secara keseluruhan, jumlah data yang berhasil dikumpulkan mencapai 2.800 file audio. Namun, agar penelitian ini lebih fokus serta untuk menjaga keseimbangan distribusi data, hanya 2.000 file audio yang dipilih untuk digunakan pada tahap pelatihan dan pengujian model. Pemangkasan data ini juga bertujuan untuk menghindari bias data yang terlalu besar pada kelas tertentu serta mempercepat proses komputasi. Setiap file audio telah diberikan label emosi sesuai kategori yang ditentukan, misalnya “happy”, “sad”, atau “disgust”. Labelisasi ini penting agar model klasifikasi dapat mempelajari pola dari data latih secara tepat. Distribusi data dalam tiap kategori emosi dibuat relatif seimbang, meskipun terdapat sedikit variasi jumlah pada masing-masing kelas. Tabel distribusi data ditampilkan untuk memperlihatkan bagaimana data tersebar pada tiap kategori emosi. Dengan adanya tabel ini, dapat diketahui bahwa dataset yang digunakan memang representatif untuk penelitian klasifikasi emosi berbasis suara. Selain itu, proses pengumpulan data juga memperhatikan kualitas rekaman**.** Data yang diambil adalah suara yang jelas, minim gangguan latar belakang, serta diucapkan dengan artikulasi yang baik agar ciri emosionalnya tetap terdengar. Hal ini penting karena kualitas rekaman yang buruk dapat memengaruhi akurasi ekstraksi fitur dan pada akhirnya menurunkan performa model klasifikasi.

Distribusi data dalam masing-masing kategori emosi dapat dilihat pada tabel berikut:

Tabel 4.1. Distribusi data

Label Emosi	Jumlah Data
Takut	400
Marah	400
Senang	400
Sedih	400
Jijik	400
Total	2000

b. Preprocessing Data

Tahapan preprocessing bertujuan untuk menyiapkan data suara agar dapat digunakan secara efektif dalam proses klasifikasi. Dari total 2.800 data yang tersedia, hanya 2.000 data yang berhasil lolos tahap preprocessing dan siap digunakan. Langkah-langkah preprocessing yang dilakukan dalam penelitian ini meliputi:

1. Noise Reduction (Pengurangan Noise)

   Suara sering kali mengandung gangguan latar belakang seperti hembusan angin, suara lingkungan, atau klik mikrofon. Proses pengurangan noise dilakukan menggunakan teknik filtering agar hanya sinyal utama (suara manusia) yang dipertahankan. Dengan demikian, ciri-ciri emosi dapat lebih jelas diekstraksi.

2. Stretching (Penyesuaian Durasi)

   Tidak semua data suara berdurasi sama. Oleh karena itu, dilakukan stretching untuk menyesuaikan panjang rekaman suara. Teknik ini mempertahankan karakteristik suara tanpa mengubah frekuensi dasarnya, sehingga informasi emosional tetap terjaga.

3. Shifting (Penggeseran Sinyal Suara)

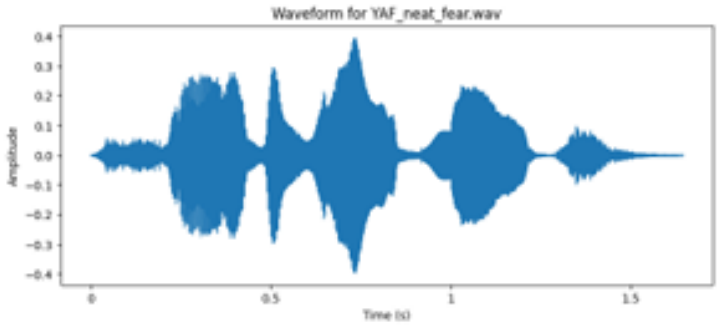
   Beberapa rekaman suara mungkin tidak dimulai tepat di awal file. Untuk mengatasi hal ini, dilakukan shifting agar sinyal suara berada pada posisi yang seimbang. Tujuannya adalah agar model tidak terkecoh oleh jeda atau keheningan di awal rekaman.

4. Pitch Adjustment (Penyesuaian Pitch)

   Pitch atau tinggi nada suara bervariasi antar individu. Agar lebih seragam, dilakukan penyesuaian pitch sehingga data suara berada dalam rentang frekuensi yang

stabil. Hal ini membantu model dalam menemukan pola umum dari setiap emosi. Hasil dari preprocessing ini dapat divisualisasikan dalam bentuk grafik sinyal suara (waveform). Grafik menunjukkan bagaimana sinyal suara telah lebih bersih, stabil, dan siap digunakan untuk tahap ekstraksi fitur berikutnya. Berikut merupakan salah satu hasil sinyal suara yang diuji coba pada penelitian ini setelah melalui tahap preprocessing.

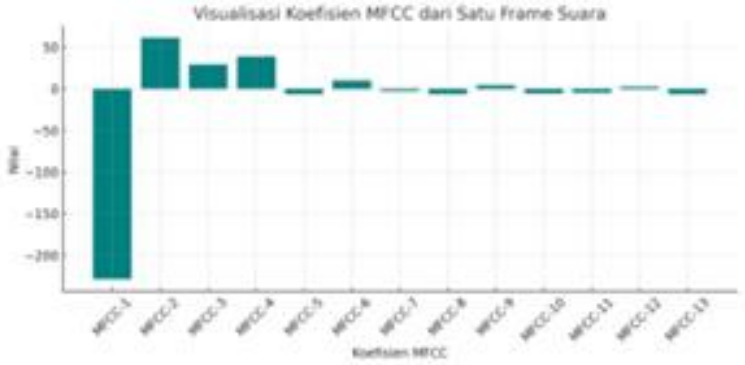
1 Emosi Fear (Takut)



Gambar 4.1 Sinyal Suara Emosi Fear (Takut)

c. Ekstraksi Fitur MFCC

Tahap berikutnya adalah ekstraksi fitur suara**.** Pada penelitian ini digunakan metode Mel-Frequency Cepstral Coefficients (MFCC)**.** MFCC merupakan salah satu metode paling populer untuk representasi suara, terutama dalam bidang pengenalan suara dan klasifikasi emosi. Proses kerja MFCC dimulai dengan membagi sinyal suara ke dalam frame-frame kecil**.** Setiap frame biasanya berdurasi 20–40 milidetik, sehingga cukup singkat untuk dianggap stabil tetapi tetap menyimpan informasi penting. Dari setiap frame, dihitung 13 koefisien utama yang merepresentasikan karakteristik spektral dari suara. Secara sederhana, MFCC mengubah sinyal suara dari domain waktu menjadi representasi numerik di domain frekuensi yang menyerupai persepsi pendengaran manusia. Karena manusia lebih peka terhadap perbedaan frekuensi rendah dibandingkan tinggi, skala Mel digunakan untuk menyesuaikan perhitungan. Dengan demikian, hasil ekstraksi MFCC berupa sekumpulan angka (fitur) yang kemudian menjadi input bagi algoritma klasifikasi. Keunggulan MFCC adalah kemampuannya untuk menangkap perbedaan halus antar emosi, seperti intonasi naik pada emosi marah atau penurunan intensitas pada emosi sedih.



Gambar 4.6 Hasil Koefisien MFCC

d. Klasifikasi Menggunakan KNN

Setelah proses ekstraksi fitur MFCC selesai, tahap selanjutnya adalah klasifikasi emosi suara menggunakan algoritma K-Nearest Neighbor (KNN). KNN dipilih karena:

1. Sederhana – algoritma ini mudah dipahami dan diimplementasikan.
2. Efektif pada dataset menengah – cocok digunakan untuk 2.000 data yang relatif cukup besar tetapi masih dapat ditangani tanpa komputasi berlebihan.
3. Tidak memerlukan asumsi distribusi data – berbeda dengan metode statistik lain, KNN hanya bergantung pada jarak antar data.

Cara kerja KNN yaitu dengan menghitung jarak antara data uji dan data latih. Jarak yang digunakan biasanya adalah Euclidean Distance. Setelah jarak dihitung, KNN akan mencari sejumlah k tetangga terdekat. Prediksi kelas ditentukan berdasarkan mayoritas dari tetangga tersebut.

Dalam penelitian ini, dilakukan pengujian terhadap beberapa nilai k: 3, 5, 7, 9, 11, dan 13. Dari hasil pengujian:

- Akurasi meningkat seiring kenaikan nilai k dari 3 hingga 9.
- Setelah k = 9, akurasi cenderung stabil hingga k = 13.
- Nilai k = 9 dipilih karena memberikan akurasi tertinggi, yaitu 90%.

Pemilihan nilai k yang lebih besar membantu mengurangi pengaruh outlier, sehingga prediksi menjadi lebih stabil.

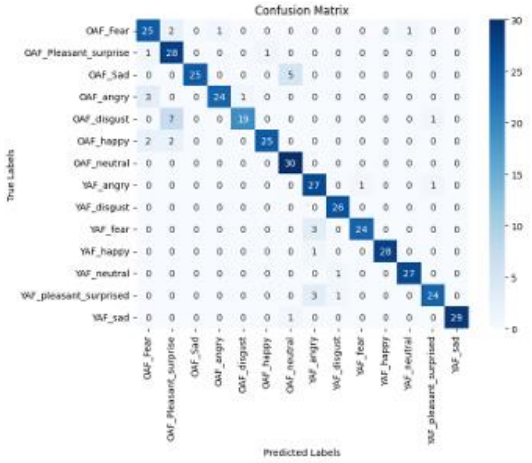
Tabel 4.2 Persentase Nilai k

Nilai k	Akurasi (%)
3	87.00
5	89.00
7	89.00
9	90.00
11	90.00
13	90.00

Dari tabel tersebut dapat dilihat bahwa akurasi meningkat dari nilai k = 3 hingga k = 9, lalu cenderung stabil hingga k = 13. Nilai k = 9 dipilih sebagai parameter terbaik karena memberikan akurasi tertinggi. Penggunaan nilai k yang lebih besar dalam konteks ini membantu meredam pengaruh data outlier dan membuat prediksi menjadi lebih stabil.

Hal ini sangat berguna untuk dataset yang relatif besar seperti 2000 data suara. Untuk melihat lebih rinci keberhasilan klasifikasi tiap kelas emosi, digunakan confusion matrix. Confusion matrix memberikan gambaran seberapa banyak data dari setiap kelas yang berhasil diprediksi dengan benar, serta seberapa banyak yang salah diklasifikasikan ke kelas lain.

Hasil visualisasi confusion matrix di sajikan dalam bentuk heatmap, yang memudahkan dalam membaca performa klasifikasi untuk masing masing kelas.



Gambar 4.7 Confusion Matrix

Untuk mengevaluasi hasil klasifikasi, digunakan confusion matrix. Matriks ini memperlihatkan jumlah data dari tiap kelas yang berhasil diprediksi dengan benar maupun salah.

Visualisasi confusion matrix ditampilkan dalam bentuk heatmap, sehingga performa klasifikasi antar kelas lebih mudah dipahami. Dari hasil penelitian diperoleh temuan sebagai berikut:

- Kelas OAF_neutral (Older Adult Female – netral) berhasil dikenali sempurna (30/30).
- Kelas YAF_sad (Younger Adult Female – sedih) juga menunjukkan performa sangat baik dengan 29/30 data dikenali dengan benar.
- Kelas YAF_happy (senang) hampir sempurna dengan 28/30 prediksi benar.

Meskipun demikian, terdapat kelas emosi tertentu yang cenderung lebih sulit dibedakan, misalnya antara takut dan jijik, karena keduanya memiliki pola akustik yang mirip.

Selain itu, metrik evaluasi lain seperti recall dan f1-score juga dihitung. Misalnya:

- Emosi YAF_happy memiliki recall sebesar 0.97 dan f1-score sebesar 0.93.
- Emosi YAF_disgust bahkan mencapai recall sempurna 1.00 dengan f1-score 0.91.

Hasil ini membuktikan bahwa model tidak hanya akurat secara keseluruhan, tetapi juga konsisten dalam mengenali berbagai kategori emosi suara.

K-Nearest Neighbors Results:
Accuracy: 0.90

Classification Report:

	precision	recall	f1-score	support
OAF_Fear	0.83	0.93	0.88	27
OAF_Pleasant_surprise	0.77	0.63	0.69	27
OAF_Sad	0.96	0.82	0.88	28
OAF_angry	0.82	1.00	0.90	28
OAF_disgust	1.00	0.72	0.84	29
OAF_happy	0.81	0.93	0.87	28
OAF_neutral	0.84	0.94	0.89	33
YAF_angry	1.00	0.77	0.87	30
YAF_disgust	0.90	0.96	0.93	27
YAF_fear	0.84	1.00	0.91	31
YAF_happy	1.00	1.00	1.00	33
YAF_neutral	0.93	0.93	0.93	30
YAF_pleasant_surprised	1.00	0.97	0.98	29
YAF_sad	0.97	0.97	0.97	31

Gambar 4.8 Hasil akhir

Dari gambar di atas secara keseluruhan, kombinasi metode MFCC untuk ekstraksi fitur dan algoritma KNN untuk klasifikasi terbukti mampu memberikan hasil yang akurat dalam deteksi emosi suara. Dengan pemilihan nilai k = 9, model berhasil mencapai tingkat akurasi terbaik sebesar 90%, serta menunjukkan performa klasifikasi yang stabil antar kelas. Emosi

YAF_happy memiliki nilai recall sebesar 0.97, yang berarti 97% data yang sebenarnya termasuk ke dalam emosi ini berhasil dikenali dengan benar oleh model, dengan f1-score sebesar 0.93 yang mencerminkan keseimbangan antara ketepatan dan kelengkapan dalam klasifikasi. Sementara itu, emosi YAF_disgust memiliki recall sempurna sebesar 1.00, yang berarti seluruh data diklasifikasikan dengan benar, dan f1-score sebesar 0.91 yang menunjukkan performa model tetap sangat baik meskipun masih terdapat sedikit kesalahan prediksi. Hasil ini menunjukkan bahwa model tidak hanya akurat secara keseluruhan, tetapi juga konsisten dalam mengenali berbagai kategori emosi suara.

Pembahasan

Berdasarkan hasil penelitian, kombinasi metode MFCC untuk ekstraksi fitur dan algoritma KNN untuk klasifikasi terbukti mampu mendeteksi emosi suara dengan baik. Dengan pemilihan parameter k = 9, model berhasil mencapai akurasi tertinggi sebesar 90%. Pemilihan nilai k ini tidak sembarangan. Jika nilai k terlalu kecil (misalnya k = 1 atau 3), model akan sangat sensitif terhadap noise dan data outlier. Sebaliknya, jika k terlalu besar, prediksi menjadi lebih stabil tetapi kehilangan sensitivitas terhadap variasi data. Oleh karena itu, k = 9 dianggap paling optimal.

Dalam hasil klasifikasi, emosi netral, senang, dan sedih lebih mudah dikenali karena memiliki ciri suara yang jelas dan konsisten. Emosi takut dan jijik seringkali tertukar karena memiliki karakteristik akustik yang mirip, misalnya intonasi menurun dengan durasi yang hampir sama. Secara umum, sistem yang dibangun sudah cukup efektif. Namun, terdapat beberapa peluang pengembangan lebih lanjut, antara lain:

1. Menambahkan fitur lain selain MFCC, misalnya Spectral Centroid atau Chroma Features, agar model memiliki representasi lebih kaya.
2. Menggunakan algoritma klasifikasi lain seperti SVM, Random Forest, atau bahkan deep learning (CNN/LSTM) untuk membandingkan performa.
3. Memperbesar dataset dengan variasi suara dari berbagai usia, jenis kelamin, dan aksen untuk meningkatkan kemampuan generalisasi model.

Kesimpulannya, penelitian ini memberikan bukti kuat bahwa KNN berbasis MFCC dapat menjadi solusi sederhana namun efektif untuk deteksi emosi berbasis suara, dengan akurasi yang cukup tinggi dan performa stabil antar kelas emosi.

KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan bahwa proses deteksi emosi berbasis suara mampu berjalan secara efektif dengan mengombinasikan metode K-Nearest Neighbors (KNN) dan ekstraksi fitur Mel Frequency Cepstral Coefficient (MFCC). MFCC memiliki peranan penting dalam tahap awal, yaitu mengubah karakteristik sinyal suara menjadi bentuk numerik yang merepresentasikan ciri khas dari setiap emosi. Representasi numerik inilah yang kemudian digunakan sebagai input untuk algoritma klasifikasi. Selanjutnya, metode KNN berfungsi dalam proses pengenalan emosi dengan cara menghitung kedekatan antara data uji dan data latih. Dengan prinsip sederhana namun efektif, KNN mampu memprediksi kelas emosi berdasarkan mayoritas tetangga terdekat. Pengujian yang dilakukan dengan berbagai variasi nilai k menunjukkan bahwa nilai k = 9 adalah parameter terbaik, karena mampu menghasilkan akurasi tertinggi, yaitu sebesar 90%. Dari hasil klasifikasi, emosi netral, senang, dan sedih cenderung lebih mudah dikenali, sebab pola suara pada kategori tersebut relatif jelas, konsisten, dan stabil. Sebaliknya, emosi seperti takut dan jijik lebih sering tertukar karena memiliki pola akustik yang mirip. Hal ini menegaskan bahwa meskipun sistem telah efektif, masih terdapat ruang pengembangan, baik melalui penambahan fitur suara lain maupun penggunaan algoritma yang lebih kompleks agar akurasi dan generalisasi model semakin meningkat.

DAFTAR PUSTAKA

Y. Khoirotul Aini, T. Budi Santoso, and T. Dutono, “Pemodelan CNN Untuk Deteksi Emosi Berbasis Speech Bahasa Indonesia,” *J. Komput. Terap.*, vol. 7, no. Vol. 7 No. 1 (2021), pp. 143–152, 2021, doi: 10.35143/jkt.v7i1.4623.

S. D. Reakaa and J. Haritha, “Comparison study on speech emotion prediction using machine learning,” *J. Phys. Conf. Ser.*, vol. 1921, no. 1, 2021, doi: 10.1088/1742-6596/1921/1/012017.

H. Judul, D. Oleh, R. Galang, and J. Respati, “Identifikasi Emosi Melalui Suara Menggunakan Support Vector Machine Dan Convolutional Neural Network,” *Tek. Inform.*, 2021.

D. Ardiyansyah and Jayanta, "Model Klasifikasi Emosi Berdasarkan Suara Manusia Dengan Metoode Multilater Perceptron," *Semin. Nas. Mhs. Ilmu Komput. dan Apl. Jakarta-Indonesia*, no. April, pp. 689–702, 2021, [Online]. Available: https://conference.upnvj.ac.id/index.php/senamika/article/view/1401

P. Widya Eka Safitri, A. Eka Karyawati, and K. Selatan, "Kombinasi Metode MFCC dan KNN dalam Pengenalan Emosi Manusia Melalui Ucapan," *Jnatia*, vol. 1, no. 1, pp. 133–140, 2022.