Perbandingan Akurasi Deteksi Emosi Pada Suara Menggunakan Multilayer Perceptron, Random Forest, Decision Tree dan K-NN

Windra Swastika[1], Alvin A. Oepojo[1], dan Paulus L. T. Irawan[1]
[1]Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Ma Chung, Malang, Indonesia

Corresponding author: Windra Swastika (e-mail: windra.swastika@machung.ac.id).

ABSTRACT This study aims to compare the accuracy of emotion recognition through sound using several types of classifiers. The four basic emotions to be recognized are happy, sad, neutral, and angry. The research methodology began with obtaining sound datasets from the RAVDESS database, consisting of 24 actors with 60 sound samples per actor. However, only 28 sounds were selected from each actor, resulting in a total of 672 sounds used in this study. Three techniques, namely mel frequency cepstral coefficient (MFCC), Chroma, and Mel Scale, were used to extract features from the sound dataset. Four types of classifiers, Multilayer Perceptron Classifier (MLPC), Decision Tree, Random Forest, and K-NN, were then used to create the models. The dataset was divided into training and testing data for each classifier in three trials, which were 85% training – 25% testing, 80% training – 25% testing, and 75% training – 25% testing. The results of the study showed that the model using the Random Forest Classifier had the highest accuracy, which was 79% with an 80% training – 20% testing dataset division. Meanwhile, the model using the Decision Tree Classifier had the lowest accuracy, which was 57% with a 75% training – 25% testing dataset division. In this study, the feature extraction techniques used, which were MFCC, Chroma, and Mel Scale, were proven effective in producing sound dataset features. Furthermore, the study results also showed that the Random Forest Classifier was superior in recognizing emotions through sound compared to other types of classifiers.

KEYWORDS Decision Tree, K-NN, Multilayer Perceptron, Random Forest, Speech Emotion Recognition

ABSTRAK Penelitian ini bertujuan untuk membandingkan akurasi pengenalan emosi melalui suara dengan menggunakan beberapa jenis classifier. Emosi dasar yang akan dikenali ada 4, yaitu senang, sedih, neutral dan marah. Metodologi penelitian dimulai dengan memperoleh dataset suara dari database RAVDESS, yang terdiri dari 24 aktor dengan jumlah suara sebanyak 60 per aktor. Namun, hanya 28 suara yang dipilih dari setiap aktor, sehingga total ada 672 suara yang digunakan dalam penelitian ini. Untuk mengekstraksi fitur dari dataset suara, digunakan tiga teknik yaitu mel frequency cepstral coefficient (MFCC), Chroma, dan Skala Mel. Kemudian, empat jenis classifier digunakan dalam pembuatan model yaitu Multilayer Perceptron Classifier (MLPC), Decision Tree, Random Forest, dan K-NN. Dataset dibagi menjadi data train dan data test dalam 3 uji coba untuk masing-masing classifier, yait 85% train – 25% test, 80% train – 25% test, dan 75% train dan 25% test. Hasil penelitian menunjukkan bahwa model dengan menggunakan Random Forest Classifier memiliki akurasi tertinggi yaitu sebesar 79% dengan pembagian dataset 80% train - 20% test. Sedangkan, model dengan Decision Tree Classifier memiliki akurasi terendah sebesar 57% dengan pembagian dataset menjadi 75% train - 25% test. Dalam penelitian ini, teknik ekstraksi fitur yang digunakan yaitu MFCC, Chroma, dan Skala Mel, yang terbukti efektif dalam menghasilkan fitur dari dataset suara. Selain itu, hasil penelitian juga menunjukkan bahwa Random Forest Classifier lebih unggul dalam mengenali emosi melalui suara jika dibandingkan dengan jenis classifier yang lain.

KATA KUNCI Decision tree, Deteksi Emosi Pada Suara, K-NN, Multilayer Perceptron, Random Forest

I. PENDAHULUAN

Emosi merupakan fenomena yang tidak dapat dipisahkan dari kehidupan manusia. Emosi dapat muncul secara verbal maupun nonverbal, dan salah satu bentuk ekspresi emosi yang paling umum adalah melalui suara. Suara yang dipenuhi emosi dapat dengan mudah diidentifikasi oleh orang lain, terutama jika suara tersebut mengalami perubahan yang signifikan dari suara normal. Deteksi emosi suara merupakan salah satu bidang yang sedang berkembang dalam penelitian teknologi suara.

Deteksi emosi suara memiliki berbagai aplikasi praktis dalam kehidupan sehari-hari [1]. Salah satu aplikasi yang paling umum adalah dalam industri telekomunikasi, di mana teknologi deteksi emosi suara dapat digunakan untuk meningkatkan layanan pelanggan. Selain itu, deteksi emosi suara juga dapat digunakan dalam bidang kesehatan mental untuk mengidentifikasi perubahan emosi yang mungkin merupakan tanda-tanda awal dari masalah kesehatan mental. Aplikasi lain dari deteksi emosi suara adalah dalam pembuatan film, di mana teknologi ini dapat digunakan untuk membantu menciptakan suasana yang tepat sesuai dengan skenario.

Meskipun deteksi emosi suara memiliki banyak aplikasi praktis, namun teknologi ini juga merupakan tantangan yang cukup sulit. Emosi manusia bisa bervariasi secara cepat dan tidak terlalu terprediksi, sehingga sulit untuk mengidentifikasi emosi dengan tepat hanya dengan menggunakan suara saja. Selain itu, ada banyak faktor yang bisa mempengaruhi emosi, seperti latar belakang, pengalaman, dan kondisi fisik, yang semuanya harus dipertimbangkan dalam proses deteksi emosi suara.

Penelitian terdahulu telah menggunakan berbagai metode untuk mencoba meningkatkan akurasi deteksi emosi suara. Salah satu metode yang sering digunakan adalah pemodelan suara, di mana algoritma dibuat untuk mengenali pola-pola suara yang biasa muncul saat seseorang merasa cemas, sedih, marah, atau senang. Selain itu, pemodelan gerakan tubuh juga sering digunakan sebagai tambahan untuk membantu dalam proses deteksi emosi suara [2]. Meskipun metode-metode ini telah menunjukkan hasil yang cukup baik, namun masih terdapat banyak tantangan yang harus diatasi untuk meningkatkan akurasi deteksi emosi suara.

Salah satu tantangan utama dalam deteksi emosi suara adalah perbedaan individu dalam ekspresi emosi [3]. Setiap orang memiliki cara yang berbeda dalam mengekspresikan emosi, sehingga sulit untuk menentukan pola-pola suara yang pasti akan muncul saat seseorang merasa cemas, sedih, marah, atau senang. Selain itu, ada juga perbedaan kultural dalam ekspresi emosi yang perlu dipertimbangkan dalam proses deteksi emosi suara. Penelitian terdahulu telah menunjukkan bahwa teknologi deteksi emosi suara masih belum cukup akurat untuk digunakan secara luas. Namun, dengan terus melakukan penelitian dan pengembangan, diharapkan teknologi deteksi emosi suara akan semakin maju dan dapat memberikan manfaat yang lebih besar bagi kehidupan sehari-hari.

Pada penelitian ini, akan dilakukan perbandingkan akurasi deteksi emosi pada suara dengan menggunakan 4 metode classifier, yaitu Multilayer Perceptron [4], Random Forest [5], Decision Tree [6] dan KNN. Metode-metode tersebut juga sering digunakan di penelitian lain untuk menyelesaikan permasalahan klasifikasi [7][8]. Dari ke-4 metode tersebut diharapkan bisa diketahui metode manakah yang dapat menghasilkan model dengan akurasi yang paling baik dalam melakukan deteksi emosi pada suara serta parameter yang sesuai untuk mendapatkan akurasi yang terbaik.

II. TINJAUAN PUSTAKA

A. MFCC

MFCC (*Mel Frequency Cepstral Coefficient*) merupakan suatu cara ekstraksi fitur yang biasanya digunakan untuk *speech recognition* [9]. Guna MFCC adalah sebagai bentuk saluran vokal menunjukkan dirinya dalam amplop spektrum daya waktu singkat. Menghitung MFCC memiliki 6 tahapan yaitu:

1. Melakukan pre-emphasis pada sinyal suara. Pre-emphasis adalah proses meningkatkan amplitudo sinyal suara pada frekuensi tinggi untuk mengurangi distorsi harmonik.
2. Menyamakan sinyal suara dengan menggunakan Fast Fourier Transform (FFT) untuk mengkonversi dari domain waktu ke domain frekuensi.
3. Menentukan filter bank mel-frequency dengan menentukan beberapa titik pemisah pada skala mel dan mengaplikasikan triangular filter pada setiap titik pemisah tersebut.
4. Melakukan dot-product antara hasil FFT dengan setiap filter dalam filter bank, yang akan memberikan dengan log-energi pada setiap filter.
5. Mengaplikasikan Discrete Cosine Transform (DCT) pada log-energi untuk mengkonversi ke domain cepstrum.
6. Koefisien cepstral dari hasil DCT yang dianggap penting akan dipilih sebagai fitur yang digunakan untuk klasifikasi suara.

B. CHROMA

Chroma adalah salah satu metode untuk mengekstraksi fitur suara yang digunakan dalam pengolahan suara dan pengenalan suara [10]. Chroma adalah representasi musik dari sinyal suara yang memfokuskan pada distribusi energi frekuensi dari sinyal suara pada setiap nada musik.

Cara untuk menghitung Chroma adalah:

1. Melakukan analisis Short-Time Fourier Transform (STFT) pada sinyal suara untuk mengkonversi dari domain waktu ke domain frekuensi.

2. Membuat filter bank yang terdiri dari 12 filter yang mewakili setiap nada musik dalam skala chroma.
3. Melakukan dot-product antara hasil STFT dengan setiap filter dalam filter bank, yang akan memberikan kita dengan energi pada setiap filter.
4. Mapping dari energi filter bank kedalam notasi musik untuk menentukan distribusi energi frekuensi pada setiap nada musik dari sinyal suara.

Chroma feature digunakan dalam musik analysis, pengenalan musik, and content-based retrieval. Salah satu keunggulan Chroma feature adalah dapat memisahkan suara dari berbagai macam instrumen musik yang memainkan notasi yang sama. Beberapa Chroma-based feature yang umum digunakan diantaranya Chroma Vector, Chroma Energy Normalized, Chroma Deviation.

C. SKALA MEL

Skala Mel adalah skala yang digunakan dalam proses transformasi frekuensi pada pengolahan suara [11]. Skala Mel adalah skala logaritmik yang lebih cocok dengan cara kerja pendengaran manusia. Pada skala Mel, perbedaan frekuensi yang sama akan diwakili oleh jarak yang sama pada skala Mel, sementara pada skala linear, perbedaan frekuensi yang sama akan diwakili oleh jarak yang berbeda.

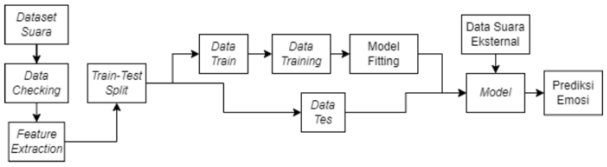
Dalam pengolahan suara, skala Mel digunakan dalam *mel frequency cepstral coefficient* (MFCC) untuk mengekstrak fitur suara. Dalam MFCC, sinyal suara dikonversi dari domain waktu ke domain frekuensi menggunakan *Fast Fourier Transform* (FFT), kemudian dikonversi lagi dari domain frekuensi linear ke domain frekuensi Mel menggunakan filter bank mel-frequency.

Pengkonversian dari domain frekuensi linear ke domain Mel dilakukan dengan melakukan transformasi (1).

$$f(mel) = 2595\log(1 + f(Hz)/700) \qquad (1)$$

III. METODOLOGI PENELITIAN

A. RANCANGAN SISTEM



GAMBAR 1. Rancangan sistem penelitian

Gambar 1 menunjukkan rancangan sistem yang akan dibuat. Hal pertama yang perlu disiapkan adalah dataset suara. Dataset suara berisi berbagai macam suara aktor dan / atau orang dengan gender yang berbeda yang berbicara dengan berbagai macam dialog dengan bahasa yang sama. Format dari *file* suara berupa .wav atau setidaknya memiliki format yang sama dan kualitas tiap *file* suara harus mono. Tiap *file* juga harus memiliki nama yang berbeda dengan format penamaan tertentu. Contohnya bila suara emosi neutral adalah 01 dengan actor 01 yang berbicara, maka file bisa dinamai 01-01. Jumlah suara harus minimal ratusan agar saat model dilatih akan memiliki banyak referensi untuk membedakan emosi.

Lalu, dataset suara tersebut akan dicek di *Data Checking*. *Data Checking* adalah sebuah tahapan untuk memastikan data bisa dipakai. Hal yang dicek seperti apakah ciri-ciri untuk tahapan selanjutnya *feature extraction* bisa diambil atau tidak dari *file* suara tersebut dan apakah tiap file memiliki kualitas *mono*. Bila ada beberapa file yang tidak bisa diambil ciri-cirinya, maka *file* tersebut tidak akan digunakan untuk tahap selanjutnya. Setelah dataset sudah dicek, tahap selanjutnya adalah *feature extraction* dimana data dilihat ciri-cirinya untuk membedakan tiap emosi. Ciri-cirinya bisa merupakan frekuensi dari suara, penamaan *file* dan lain-lainnya.

Lalu data akan dibagi dengan *train-test split* untuk memisahkan dataset suara menjadi dua data yaitu data latih untuk melatih model yang akan digunakan untuk prediksi dan data tes untuk mengetes prediksi dari data latih. Pembagian data latih dan data tes bisa dipisah sesuai keinginan. Pada penelitian ini, akan dilakukan 3 jenis *train-test split*, yaitu *85% train – 15% test, 80% train – 20% test* dan *75% train – 25% test.*

Sebelum data bisa dimasukkan, model harus dibuat dahulu. Model ini berisi sebuah algoritma yang digunakan untuk membedakan data-data menjadi kategori-kategori tertentu seperti membedakan emosi senang, sedih, neutral dan marah. Model juga diatur sesuai algoritmanya seperti jumlah iterasi, jumlah *hidden layer,* dan berbagai macam pengaturan lainnya. Untuk penelitian ini, ada empat algoritma model yang akan digunakan yaitu MLPClassifier (*Multi-Layer Perceptron Classifier*), *Decision Tree Classifier*, *Random Forest Classifier*, dan *K-Nearest Neighbors Classifier*.

Bila data sudah dipisah, maka data latih dilatih dengan mempelajari tiap file yang memililki fitur yang berbeda dan penamaan file yang berbeda dari emosi tertentu. Akhirnya, data dimasukkan ke dalam sebuah model yaitu sebuah file yang sudah dilatih untuk mengerti perbedaan dari tiap emosi untuk melakukan prediksi dengan data tes. Prediksi dilakukan dengan membandingkan hasil latihan yang akan dibandingkan dengan emosi sebenarnya. Akurasi dari model akan tergantung dari total jumlah benar dan salah dari prediksi model.

B. FEATURE EXTRACTION

Feature extraction berguna untuk melihat ciri-ciri / fitur dari sebuah file untuk membedakan perbedaan dari satu file dengan yang lain sesuai yang diinginkan. Ciri-ciri atau fitur yang dibedakan ada empat:

1. Nama file,
2. MFCC (*Mel Frequency Cepstral Coefficient*) yaitu spektrum daya suara jangka pendek,
3. Chroma yang berkaitan dengan 12 kelas *pitch* yang berbeda,

4. Mel (*Mel Spectrogram Frequency*) untuk mengetahui frekuensi suara.

File suara akan dicek fitur atau ciri suara yang dijelaskan di atas ada atau tidak kecuali nama file dan bila ada, akan mengambil beberapa hal untuk membantu membedakan fitur-fitur satu file dengan file lain. MFCC akan mengambil *sample rate* dan jumlah MFCC dari file suara, Chroma mengambil STFT (*Short-time Fourier transform*) dan *sample rate* dan Mel hanya memerlukan *sample rate*. Hasil dari tiap fitur akan direpresentasikan dengan satuan angka.

Setelah tiap ciri atau fitur sudah didapat, maka tiap fitur dan / atau ciri-ciri akan disimpan dalam sebuah array untuk membedakan tiap file yang ada. Dalam penelitian ini, hal yang ingin dibedakan adalah emosi. Ada berbagai macam emosi dan juga emosi memiliki cara mengekspresikannya yang berbeda seperti intensitas kemarahan, intensitas kesenangan dan lain-lain berbasis dari empat fitur diatas.

Dari banyak emosi yang bisa diprediksi, penelitian ini akan menggunakan empat emosi umum yaitu neutral, senang, sedih dan marah. Untuk mengetahui bagaimana mengerti emosi yang berbeda, maka perlu membuat *dictionary* untuk emosi tersebut. *Dictionary* dibuat dengan membuat array dimana string di nama file akan menunjukkan emosi tertentu, contohnya seperti '01' menunjukkan emosi neutral. Setelah itu, file suara mulai diekstrak satu per satu untuk tiap aktor dan tiap emosi yang ada.

C. DATASET SUARA

Untuk penelitian ini, ada dua dataset suara yang akan digunakan. Dataset suara pertama berasal dari website Kaggle dengan dataset RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song) Emotional Speech Audio [12] yang memiliki 24 aktor (12 laki-laki dan 12 perempuan) dan setiap aktor memiliki 60 suara tetapi hanya 28 suara dari tiap aktor yang dipakai dengan total 672.

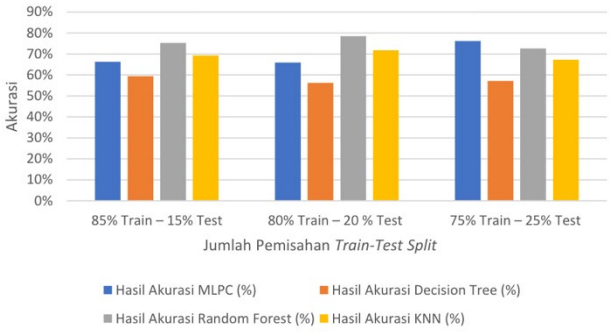
Ada tambahan 8 file suara yang berasal dari *website* Freesound dengan 4 suara dari aktor laki-laki yang berbeda dan 4 suara aktor perempuan yang berbeda-beda yang digunakan untuk mengetes model dengan data eksternal. Penamaan dari data suara eksternal tidak disamakan dengan format nama dari dataset RAVDESS.

Dataset RAVDESS memiliki penamaan spesifik untuk tiap *file* dan dikode dengan nomor. Berikut adalah ciri-ciri dari penamaan file:

1. Modalitas (01 = *video* dan *audio*, 02 = *video* saja, 03 = *audio* saja),
2. Lagu atau monolog (01 = monolog, 02 = lagu).
3. Emosi (01 = *neutral*, 02 = tenang, 03 = senang, 04 = sedih, 05 = marah, 06 = ketakutan, 07 = merasa jijik, 08 = kaget),
4. Intensitas emosi (01 = normal, 02 = kuat). Catatan: Neutral tidak memiliki intensitas emosi,
5. Kata (01 = "*Kids are talking by the door*", 02 = "*Dogs are sitting by the door*", 03 = Lain-lain 1, 04 = Lain-lain 2),
6. Repetisi (01 = repetisi pertama, 02 = repetisi kedua),
7. Aktor (Nomor ganjil adalah laki-laki, nomor genap adalah perempuan).

IV. HASIL DAN DISKUSI

Model akan dibuat dengan berbagai macam classifier yaitu MLPC, *Decision Tree Classifier*, *Random Forest Classifier* dan KNN *Classifier*.



GAMBAR 2. Hasil akurasi untuk masing-masing model

Gambar 2 menunjukkan hasil akurasi dari 4 *classifier* dan masing-masing classifier diukur menggunakan 3 jenis kombinasi *train-test*.

Rata-rata akurasi dari ke-4 *classifier* adalah sebagai berikut: untuk 85% train – 15% test adalah 67.6%, rata-rata akurasi untuk 80% train – 20% test adalah 68.2% sedangkan untuk 75% train - 25% test memiliki rata-rata akurasi 68.3%. Sedangkan untuk masing-masing classifier, rata-rata akurasi yang didapatkan adalah sebagai berikut: untuk MLPC, rata-rata akurasi yang didapatkan adalah 69.5%, untuk rata-rata akurasi yang didapatkan dari *Decision Tree* adalah 57.6%, untuk rata-rata akurasi yang didapatkan dari *Random Forest* adalah 75.5%, dan rata-rata akurasi yang didapatkan KNN adalah 68.2%.

Secara umum, hasil akurasi tertinggi didapatkan untuk jenis 80% train - 20% test dengan *Random Forest Classifier* yaitu dengan akurasi sebesar 79%. Sementara akurasi terendah adalah model dengan 75% train - 25% test yang menggunakan *Decision Tree Classifier* yaitu dengan akurasi 57%.

Tabel I, II, III dan IV merupakan hasil *confusion matrix* dari masing-masing classifier.

TABEL I
CONFUSION MATRIX PREDIKSI EMOSI MODEL MLPC

Emosi	Neutral (Prediksi)	Happy (Prediksi)	Sad (Prediksi)	Angry (Prediksi)
Neutral (Asli)	47.3%	30.1%	29.1%	11%
Happy (Asli)	3.1%	67.7%	17.7%	11.6%
Sad (Asli)	4.2%	18%	73%	7.3%
Angry (Asli)	0%	10.1%	5%	84.9%

TABEL II
CONFUSION MATRIX PREDIKSI EMOSI MODEL DECISION TREE CLASSIFIER

Emosi	Neutral (Prediksi)	Happy (Prediksi)	Sad (Prediksi)	Angry (Prediksi)
Neutral (Asli)	43.7%	16.4%	34.5%	5.4%
Happy (Asli)	8.5%	53.8%	13.1%	24.6%
Sad (Asli)	23.3%	21.7%	50.8%	4.2%
Angry (Asli)	2%	7.1%	2%	81.9%

TABEL III
CONFUSION MATRIX PREDIKSI EMOSI MODEL RANDOM FOREST CLASSIFIER

Emosi	Neutral (Prediksi)	Happy (Prediksi)	Sad (Prediksi)	Angry (Prediksi)
Neutral (Asli)	49.1%	12.7%	38.2%	0%
Happy (Asli)	1.5%	69.2%	17%	12.3%
Sad (Asli)	4.2%	15%	77.5%	3.3%
Angry (Asli)	0%	5%	2%	93%

TABEL IV
CONFUSION MATRIX PREDIKSI EMOSI MODEL KNN CLASSIFIER

Emosi	Neutral (Prediksi)	Happy (Prediksi)	Sad (Prediksi)	Angry (Prediksi)
Neutral (Asli)	76%	11%	13%	0%
Happy (Asli)	4%	61%	13%	22%
Sad (Asli)	17%	14%	62%	7%
Angry (Asli)	2%	7.1%	4%	86.9%

Dari tabel I, II, III dan IV, terlihat bahwa model yang prediksi emosi yang paling akurat adalah model yang menggunakan *Random Forest Classifier* dengan rata-rata akurasi 72.2%. Emosi yang paling mudah diprediksi adalah emosi *Angry* atau marah dengan akurasi di atas 93%.

V. KESIMPULAN

Dari keseluruhan model *classifier* yang diuji untuk melakukan prediksi emosi berdasarkan suara, didapatkan bahwa *Random Forest Classifier* memiliki akurasi yang tertinggi, yaitu 75,5%. Hal ini dikarenakan *classifier* ini menggabungkan beberapa pohon keputusan (*decision tree*) untuk menghasilkan prediksi akhir. Ketika beberapa pohon keputusan digabungkan, akan terbentuk "hutan" yang dapat memberikan hasil yang lebih akurat dan stabil. Dalam hal ini, *Random Forest Classifier* dapat mengatasi *overfitting* yang sering terjadi pada *decision tree*, serta dapat mengatasi masalah *high variance* yang biasa terjadi pada *MLPC* dan *K-NN*.

Selain itu, Random Forest Classifier juga dapat melakukan seleksi fitur secara otomatis dan memilih fitur-fitur yang paling relevan dalam mengklasifikasikan data. Hal ini sangat penting dalam pengenalan emosi melalui suara karena suara memiliki banyak fitur yang kompleks dan tidak semuanya relevan dalam menentukan emosi yang diungkapkan.

Dari *Confusion matrix* untuk *Random Forest Classifier*, didapatkan bahwa nampak bahwa emosi marah (*angry*) merupakan emosi yang paling akurat untuk dideteksi, sedangkan emosi netral merupakan emosi yang paling tidak akurat. Faktor jumlah dataset untuk emosi netral tidak sebanyak dibandingkan emosi yang lainnya memberikan pengaruh terhadap akurasi deteksi.

PERAN PENULIS

Windra Swastika: mendesain metodologi penelitian dan evaluasi
Alvin Andrius Oepojo: implementasi algoritma ekstraksi fitur; pengolah data dan ekstraksi fitur
Paulus Lucky Tirma Irawan: Pengujian hasil dan analisis

COPYRIGHT

DAFTAR PUSTAKA

[1] Zhang, Hongli, Alireza Jolfaei, and Mamoun Alazab. "A face emotion recognition method using convolutional neural network and image edge computing." *IEEE Access* 7 (2019): 159081-159089.
[2] Gunes, Hatice, and Maja Pantic. "Automatic, dimensional and continuous emotion recognition." *International Journal of Synthetic Emotions (IJSE)* 1.1 (2010): 68-99.
[3] Khalil, Ruhul Amin, et al. "Speech emotion recognition using deep learning techniques: A review." *IEEE Access* 7 (2019): 117327-117345.
[4] Alnuaim, A. A., Zakariah, M., Shukla, P. K., Alhadlaq, A., Hatamleh, W. A., Tarazi, H., ... & Ratna, R. (2022). Human-computer interaction for recognizing speech emotions using multilayer perceptron classifier. Journal of Healthcare Engineering, 2022.
[5] Yan, S., Ye, L., Han, S., Han, T., Li, Y., & Alasaarela, E. (2020, June). Speech interactive emotion recognition system based on random forest. In *2020 International Wireless Communications and Mobile Computing (IWCMC)* (pp. 1458-1462). IEEE.
[6] Sun, L., Fu, S., & Wang, F. (2019). Decision tree SVM model with Fisher feature selection for speech emotion recognition. *EURASIP Journal on Audio, Speech, and Music Processing*, *2019*(1), 1-14.
[7] L. Alwi, A. T. Hermawan, and Y. . Kristian, "Identifikasi Biji-Bijian Berdasarkan Ekstraksi Fitur Warna, Bentuk dan Tekstur Menggunakan Random Forest", INSYST, vol. 1, no. 2, pp. 92–98, Dec. 2019.
[8] J. A. Septian, T. M. Fachrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor", *INSYST*, vol. 1, no. 1, pp. 43–49, Aug. 2019.
[9] Zheng, Fang, Guoliang Zhang, and Zhanjiang Song. "Comparison of different implementations of MFCC." *Journal of Computer science and Technology* 16.6 (2001): 582-589.

[10] Er, Mehmet Bilal, and Ibrahim Berkan Aydilek. "Music emotion recognition by using chroma spectrogram and deep visual features." *International Journal of Computational Intelligence Systems* 12.2 (2019): 1622-1634.
[11] Gowdy, John N., and Zekeriya Tufekci. "Mel-scaled discrete wavelet coefficients for speech recognition." *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 00CH37100)*. Vol. 3. IEEE, 2000.
[12] Steven R. Livingstone, & Frank A. Russo. (2019). *RAVDESS Emotional speech audio* [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/256618