Kelompok 19:

2501963040 - Jason Hendrawan

2540120313 - Jonathan Adrian

2501962214 - Verrel Juanto Lukmana

1. Judul:

Human Emotion Recognition Using Deep Learning

1. Deskripsi Proyek:

Proyek ini bertujuan untuk mengolah data audio dengan tujuan menganalisa dan memprediksi jenis emosi yang terkandung dalam rekaman suara. Kami akan menggunakan teknik-teknik pemrosesan audio untuk mengidentifikasi dan mengklasifikasikan emosi yang terkandung dalam suara.

1. Dataset dan Penjelasannya:

Dataset yang kita gunakan berasal dari Kaggle dengan nama "Toronto Emotional Speech Set (TESS)". Dataset ini terdiri dari 2800 data audio yang telah diberi label jenis emosi, yang dibagi menjadi 7 kategori yaitu fear, pleasant surprise, sad, angry, disgust, happy dan neutral.

1. Tahap-tahap Eksperimen:

* Persiapan data:

Data path yang diperoleh diberi label dan kemudian diubah menjadi series. Selanjutnya, series tersebut dikonversi menjadi data frame, dan akhirnya data diacak (shuffle).

* Pre-processing data:

Langkah pertama dalam preprocessing adalah add noise. Tujuan dari penambahan noise adalah membuat data lebih robust dan dapat menangani kebisingan. Setelah itu, dilakukan proses stretch, shift, dan pitch scaling pada data.

* Ekstraksi fitur:

Pada tahap feature extraction, fitur-fitur dari data diambil menjadi beberapa jenis fitur, yaitu Zero Crossing Rate, Chroma Feature, MFCC, Root Mean Square Energy, dan Mel Spectrogram.

* Data split:

Pada tahap data split, data dibagi menjadi dua bagian, yaitu x dan y. Kemudian dilakukan label encoding untuk variabel y. Selanjutnya, data tersebut dibagi menjadi data train dan data test, dengan bagian x disebut sebagai xtrain dan xtest, sedangkan bagian y disebut sebagai ytrain dan ytest. Kita menggunakan StandardScaler pada xtrain dan xtest untuk melakukan penskalaan.

* Model Evaluation:

Dalam evaluasi model, terdapat beberapa metrik yang digunakan, antara lain, accuracy, loss, precision, recall, f1-score, support dan confusion matrix.

1. Fitur:

Fitur-fitur yang digunakan dalam eksperimen ini antara lain: Zero Crossing Rate, Chroma Feature, MFCC, Root Mean Square Energy, Mel Spectrogram

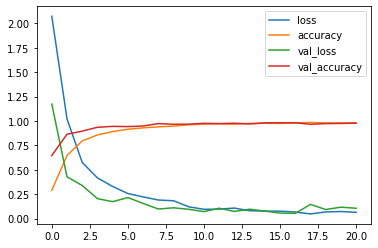
1. Metode:

Metode yang digunakan adalah CNN (Convolutional Neural Network) untuk memprediksi jenis emosi dari setiap data audio.

1. Rencana Evaluasi:

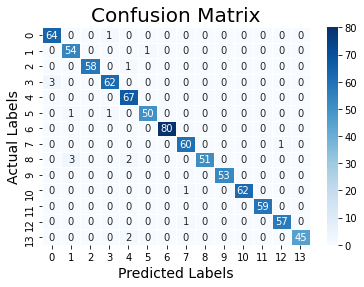
Metrik yang digunakan untuk mengevaluasi model kami adalah besar accuracy, loss, precision, recall, f1-score, support dan confusion matrix.

1. Result

****

Gambar. 1. Graphic loss, accuracy, val\_loss, dan val accuracy

Performa model kita dalam mendeteksi emosi dari data audio telah dievaluasi menggunakan dataset yang terdiri dari data pelatihan dan validasi, dan memberikan hasil sebagai berikut. Gambar. 1. menunjukkan pola pelatihan yang umum, di mana nilai loss terus berkurang hingga mencapai 0.0862 dan akurasi meningkat hingga mencapai 0.9754 seiring dengan proses pembelajaran dan peningkatan model dari epoch ke epoch. Metric validasi mengindikasikan seberapa baik model menggeneralisasi data yang belum pernah dilihat sebelumnya, dan dengan akurasi yang tinggi dan stabil sebesar 0.9786 serta nilai loss yang rendah yaitu 0.1058, hal ini menunjukkan bahwa model tersebut memiliki performa yang baik dan tidak mengalami overfitting ataupun underfitting.



Gambar. 2. Confusion matrix

Tabel 1. Hasil evaluation metrics untuk test data

| **Emotion** | **Precision** | **Recall** | **F1-score** | **Support** |
| --- | --- | --- | --- | --- |
| OAF\_Fear | 0.96 | 0.98 | 0.97 | 65 |
| OAF\_Pleasant\_surprise | 0.93 | 0.98 | 0.96 | 55 |
| OAF\_Sad | 1 | 0.98 | 0.99 | 59 |
| OAF\_Angry | 0.97 | 0.95 | 0.96 | 65 |
| OAF\_Disgust | 0.93 | 1 | 0.96 | 67 |
| OAF\_Happy | 0.98 | 0.96 | 0.97 | 52 |
| OAF\_Neutral | 1 | 1 | 1 | 80 |
| YAF\_Angry | 0.97 | 0.98 | 0.98 | 61 |
| YAF\_Disgust | 1 | 0.91 | 0.95 | 56 |
| YAF\_Fear | 1 | 1 | 1 | 53 |
| YAF\_Happy | 1 | 0.98 | 0.99 | 63 |
| YAF\_Neutral | 1 | 1 | 1 | 59 |
| YAF\_Pleasant\_surprised | 0.98 | 0.98 | 0.98 | 58 |
| YAF\_Sad | 1 | 0.96 | 0.98 | 47 |
|  | | | | |
| ACCURACY | 0.9785 | | | |
| LOSS | 0.1058 | | | |

Setelah melakukan testing terhadap model kami, kami mendapatkan hasil berikut. Tabel. 1. menampilkan hasil testing berupa precision, recall, f1-score, dan support untuk setiap label/kategori dari data pengujian. Selain itu, tabel tersebut juga mencantumkan accuracy dan loss dari hasil testing, yaitu 97,85% dan 0,1058. Accuracy dari hasil testing tersebut dapat dievaluasi juga menggunakan confusion matrix yang terdapat di Gambar. 2. Dari Gambar. 2, terlihat bahwa dari total 838 data testing, 822 data terprediksi dengan benar sesuai dengan label yang sebenarnya. Dengan demikian, dapat disimpulkan dari confusion matrix pada Gambar. 2. bahwa sebesar 98,09% data terprediksi dengan benar atau memiliki akurasi sebesar 98,09%.

1. Conclusion

Dalam project kami, kami telah membuat sebuah model untuk mendeteksi emosi seseorang dari data audio. Model yang kami buat adalah CNN yang menggunakan architecture yang kami kembangkan sendiri. Berdasarkan hasil training dan testing, dapat disimpulkan bahwa model yang telah kami buat terbukti efektif dan akurat dalam memprediksi emosi seseorang dari data audio. Dengan demikian, model kami dapat berkontribusi dalam mengoptimalkan interaksi manusia dengan teknologi, memahami dan merespons emosi manusia dengan lebih baik, serta mengambil keputusan yang lebih baik dalam berbagai aspek kehidupan. Namun, model yang kami buat masih perlu terus dikembangkan dan dilatih menggunakan berbagai dataset audio sebelum dapat diaplikasikan dalam kehidupan sehari-hari.

Notebook: <https://www.kaggle.com/code/verreljuantolukmana/emotional-speech-recognition-audio-process-all?scriptVersionId=134396246>