

Klasifikasi suara kebisingan Kota menggunakan Deep Learning Multi-Layer Perceptrons (MLP)

Ronaldito Juan Bantaras T*

¹ Informatika, Universitas Udayana, Badung, Indonesia

ARTICLE INFO

Article history:

Received Mei 22, 2021

Revised Mei 29, 2021

Accepted Agustus 14, 2021

Available online November 25, 2021

Kata Kunci:

3-5 Kata Kunci Dipisahkan Dengan Tanda Koma

Keywords:

Please Provide 3-5 Words Of Keywords Separated By Comas



This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

Copyright © 2022 by Author. Published by Universitas Pendidikan Ganesha.

ABSTRAK

Studi ini membahas penerapan Multi-Layer Perceptrons (MLP) dalam klasifikasi suara kota dengan memanfaatkan Mel-frequency cepstral coefficients (MFCC) sebagai ekstraksi fitur. Tujuan penelitian ini adalah untuk mengenali dan mengklasifikasikan suara-suara yang umumnya terdengar di lingkungan perkotaan. Sebanyak sembilan jenis suara yang biasa terdengar di kota telah diidentifikasi dan diklasifikasikan menggunakan pendekatan MLP. Data suara direkam dari lingkungan kota yang berbeda, kemudian diolah dengan metode MFCC untuk mengekstrak fitur-fitur penting. MLP dilatih menggunakan dataset yang telah dilabeli dengan berbagai suara kota. Hasil eksperimen menunjukkan bahwa pendekatan ini mampu mengklasifikasikan suara-suara kota dengan tingkat akurasi yang signifikan, membuktikan keefektifan metode MLP dan MFCC dalam klasifikasi suara lingkungan perkotaan.

ABSTRACT

This study discusses the application of Multi-Layer Perceptrons (MLP) in urban sound classification by utilizing Mel-frequency cepstral coefficients (MFCC) as feature extraction. The objective is to recognize and classify commonly heard urban sounds. Nine types of urban sounds have been identified and classified using the MLP approach. Sound data were recorded from different urban environments and processed using MFCC to extract important features. MLP was trained using a labeled dataset containing various urban sounds. Experimental results demonstrate that this approach is capable of classifying urban sounds with a significant level of accuracy, proving the effectiveness of MLP and MFCC methods in urban sound classification.

1. PENDAHULUAN

Suara lingkungan perkotaan merupakan sumber informasi penting yang dapat merefleksikan kondisi dan aktivitas di kota. Namun, suara perkotaan juga memiliki karakteristik yang bervariasi, tumpang tindih, dan tidak terstruktur, sehingga menimbulkan tantangan dalam mengenali dan mengklasifikasikannya. Penelitian klasifikasi suara perkotaan bertujuan untuk mengembangkan sistem yang dapat mengidentifikasi jenis suara yang ada dalam rekaman audio perkotaan, seperti klakson mobil, anjing menggonggong, atau musik jalanan. Klasifikasi suara perkotaan memiliki banyak manfaat, seperti untuk monitoring lingkungan, manajemen lalu lintas, keamanan publik, dan teknologi responsif.

Penelitian sebelumnya telah menggunakan berbagai metode untuk klasifikasi suara perkotaan, seperti jaringan saraf tiruan (Artificial Neural Network, ANN), jaringan saraf konvolusional (Convolutional Neural Network, CNN), jaringan saraf berulang (Recurrent Neural Network, RNN), dan mesin pendukung vektor (Support Vector Machine, SVM). Metode-metode ini biasanya membutuhkan ekstraksi fitur dari data audio, seperti koefisien cepstral frekuensi Mel (Mel-frequency cepstral coefficients, MFCC), spektrogram, atau kromagram. Fitur-fitur ini merepresentasikan karakteristik spektral dan temporal dari suara, tetapi tidak selalu mencerminkan konten semantik atau makna dari suara tersebut.

Penelitian ini mengusulkan penggunaan jaringan saraf perseptron multi-lapis (Multi-layer Perceptron, MLP) dan MFCC untuk klasifikasi suara perkotaan. MLP adalah salah satu jenis jaringan saraf tiruan yang terdiri dari beberapa lapisan neuron yang saling terhubung. MLP dapat mempelajari hubungan non-linear antara fitur masukan dan keluaran, serta dapat mengadaptasi bobotnya sesuai dengan data yang diberikan. Dengan menggunakan MLP, diharapkan dapat meningkatkan akurasi klasifikasi suara perkotaan dengan memanfaatkan kemampuan MLP dalam memodelkan kompleksitas dan variasi suara perkotaan.

Penelitian ini menggunakan dataset UrbanSound8K, yang merupakan kumpulan 8.732 rekaman audio perkotaan yang dikategorikan ke dalam 10 kelas suara. Dataset ini diambil dari berbagai sumber dan lingkungan, sehingga mencerminkan keragaman suara perkotaan yang nyata. Teknik MFCC digunakan untuk mengekstrak fitur dari data audio, yang kemudian digunakan sebagai masukan untuk MLP. MLP

*Corresponding author.

E-mail addresses: penulis1@gmail.com (Penulis Pertama)

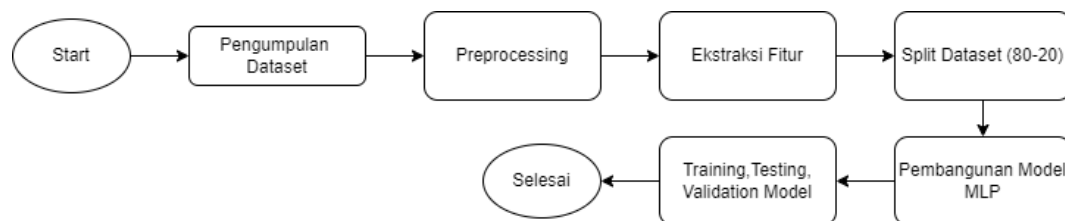
dilatih menggunakan metode optimisasi Adam dengan fungsi biaya entropi silang kategorikal. Tujuan penelitian ini adalah mengembangkan sistem klasifikasi suara perkotaan yang akurat, cepat, dan efisien, yang dapat digunakan untuk berbagai aplikasi terkait suara perkotaan.

Penelitian ini memberikan kontribusi dalam bidang pengolahan sinyal suara dan pembelajaran mendalam, dengan mengaplikasikan metode MLP untuk klasifikasi suara perkotaan. Penelitian ini juga memberikan analisis komparatif antara metode MLP dan metode lain yang telah digunakan sebelumnya, seperti CNN, RNN, dan SVM. Penelitian ini diharapkan dapat meningkatkan pemahaman tentang karakteristik dan klasifikasi suara perkotaan, serta membuka peluang untuk penelitian lanjutan di masa depan.

2. METODE

Langkah-langkah penelitian dimulai dengan eksplorasi dataset URBANSOUND8K untuk memahami distribusi kelas suara, durasi, serta karakteristik lainnya. Setelah eksplorasi data, tahap preproses dilakukan untuk membersihkan dan menyiapkan data, termasuk normalisasi dan penanganan nilai yang hilang atau tidak valid. Setelah preproses data, dilakukan teknik ekstraksi fitur menggunakan MFCC. MFCC bertujuan mengonversi sinyal suara menjadi representasi fitur yang lebih abstrak, memungkinkan analisis karakteristik frekuensi dan waktu dari suara. Library Librosa's `mfcc()` digunakan untuk menghasilkan MFCC dari data audio dalam bentuk time series.

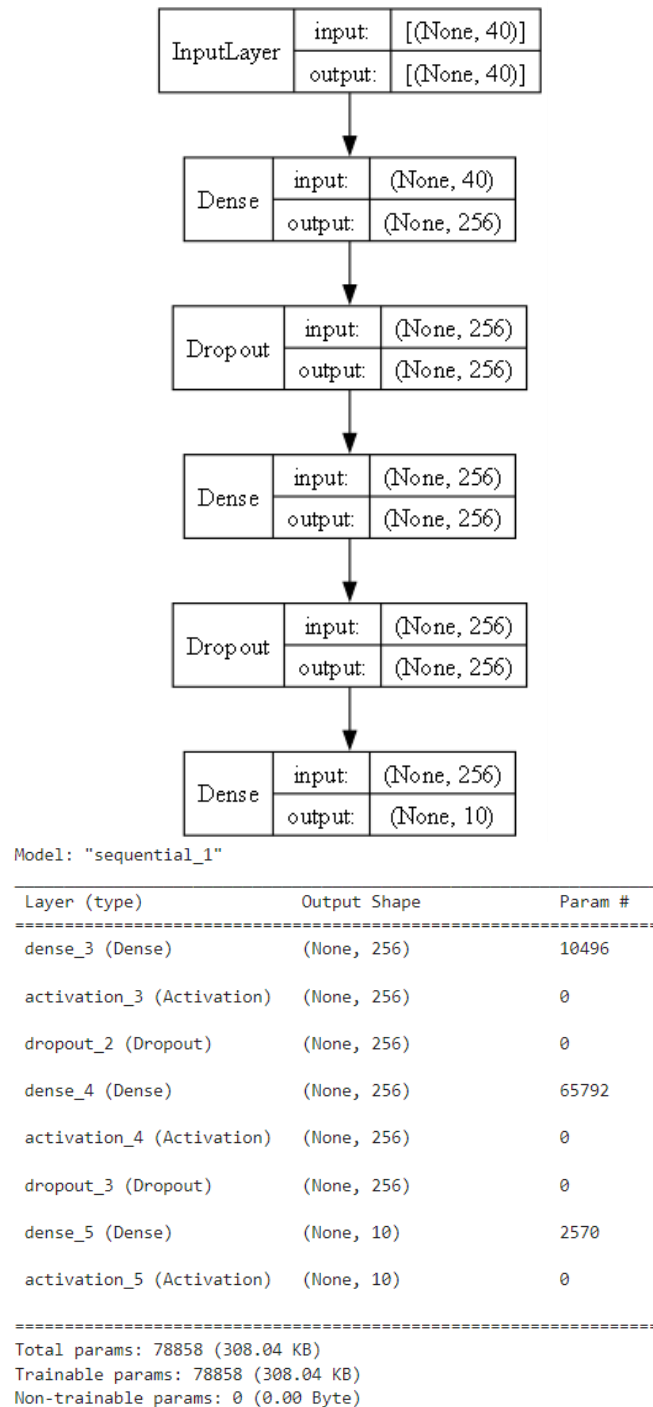
Langkah berikutnya adalah melabeli data sesuai dengan kelas suara yang tepat dan membagi dataset menjadi dua bagian dengan perbandingan 80-20, di mana 80% data digunakan untuk proses pelatihan (training) dan 20% untuk pengujian (testing). Ini memastikan model yang dibangun akan belajar dari sebagian besar data namun juga diuji pada data yang belum pernah dilihat sebelumnya.



Gambar 1. Flow Penelitian

Proses implementasi melibatkan model Multilayer Perceptron (MLP), yang dilatih, diuji, dan divalidasi. Selama proses ini, parameter dan arsitektur model disesuaikan dan dievaluasi untuk memperoleh hasil yang optimal dalam mengklasifikasikan suara bising dari lingkungan perkotaan. Dengan menggunakan model yang terlatih, prediksi dilakukan terhadap data pengujian untuk mengukur kinerja dan keakuratannya.

Penelitian ini tidak hanya melakukan analisis fitur pada data suara tetapi juga mengimplementasikan model yang dapat memahami dan mengklasifikasikan suara bising dari lingkungan perkotaan dengan akurasi yang optimal, membantu dalam pemahaman dan penanganan suara bising di lingkungan perkotaan. Implementasi melibatkan model Multilayer Perceptron (MLP), yang dilatih, diuji, divalidasi, dan kemudian digunakan untuk melakukan prediksi terhadap data pengujian. Selama proses ini, parameter dan arsitektur model diatur dan dievaluasi untuk memperoleh hasil yang optimal dalam mengklasifikasikan suara-suara bising dari lingkungan perkotaan, dan. Setelah model terlatih, langkah selanjutnya adalah memperluas manfaatnya dengan mengimplementasikannya ke dalam GUI (Graphical User Interface) menggunakan Streamlit. Hal ini memungkinkan penggunaan model secara lebih intuitif dan interaktif.



Gambar 1. Arsitektur MLP

Penjelasan Arsitektur Model MLP pada Gambar 1. adalah sebagai berikut. Lapisan masukan menerima bentuk masukan sebesar (1x40), yang sesuai dengan 40 MFCCs pada setiap sampel.

Dua lapisan pertama memiliki 256 node masing-masing dengan menggunakan fungsi aktivasi ReLU (Rectified Linear Activation), yang terbukti efektif dalam jaringan neural.

Dropout sebesar 50% diterapkan pada dua lapisan pertama. Hal ini bertujuan untuk secara acak mengabaikan sebagian node pada setiap siklus pembaharuan, sehingga memungkinkan jaringan neural memiliki generalisasi yang lebih baik dan mengurangi risiko overfitting terhadap data pelatihan.

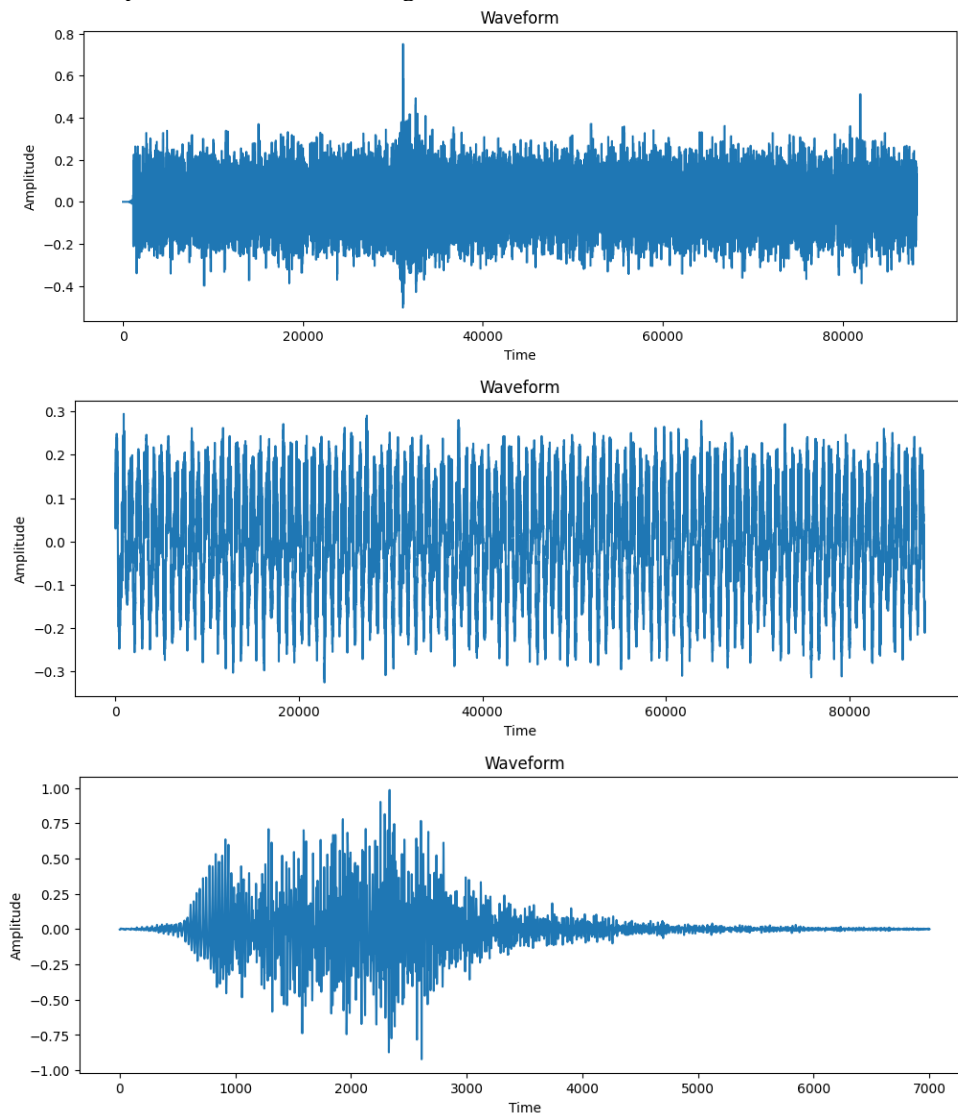
Lapisan keluaran terdiri dari 10 node (sesuai dengan jumlah klasifikasi yang mungkin), dengan fungsi aktivasi softmax. Softmax menghasilkan keluaran yang jumlahnya 1, yang dapat diinterpretasikan sebagai probabilitas. Model kemudian membuat prediksi berdasarkan opsi dengan probabilitas tertinggi.

Hasil eksperimen menunjukkan bahwa model ini memiliki potensi untuk meningkatkan klasifikasi audio berdasarkan fitur MFCCs.

3. HASIL DAN PEMBAHASAN

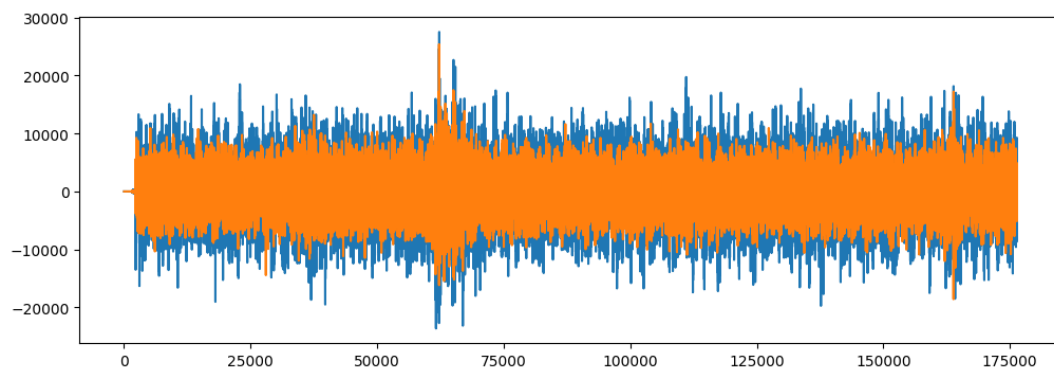
Hasil

Hasil dari Ekplorasi Data adalah sebagai berikut :



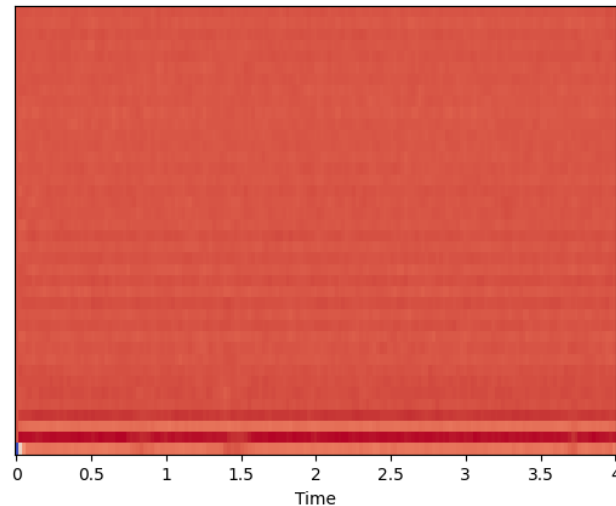
Gambar 2. Visualiasi Waveplot Dataset : klakson mobil, mesin, gonggongan anjing

Setelah melakukan Ekplorasi Data selanjutnya dilakukan Preprocessing data, dengan mengubah sample rate audio, bit-dept, dan penggabungan channel audio.



Gambar 3. Waveplot Penggabungan audio menggunakan Librosa

Setelah melakukan Preprocessing data selanjutnya dilakukan Ekstraksi Fitur, dengan menggunakan Mel-Frequency Cepstral Coefficients (MFCC) dari data sample audio yang telah di preprocessing. MFCC digunakan untuk merepresentasikan distribusi frekuensi dalam jendela waktu audio. Ini memungkinkan analisis karakteristik frekuensi dan waktu suara serta identifikasi fitur penting untuk klasifikasi. Ekstraksi fitur menggunakan MFCC pada data audio yang sudah di-preprocessing membantu dalam mendapatkan informasi krusial untuk klasifikasi suara



Gambar 4. Visualisasi librosa menghitung serangkaian 40 MFCC dalam 173 frame.

Setelah ekstraksi fitur, langkah selanjutnya adalah memberi label pada data dan membagi dataset. Kita akan menggunakan `sklearn.preprocessing.LabelEncoder` untuk mengubah data teks kategorikal menjadi data numerik yang dapat dimengerti oleh model.

Setelah proses label encoding, kita akan menggunakan `sklearn.model_selection.train_test_split` untuk membagi dataset menjadi set pelatihan dan pengujian. Ukuran set pengujian akan menjadi 20%, dan kita akan menetapkan sebuah nilai random state untuk memastikan hasil yang dapat direproduksi.

Setelah tahap pembagian dataset, langkah selanjutnya adalah membangun, melatih, mengevaluasi Jaringan Saraf Tiruan (MLP) menggunakan Keras dengan backend Tensorflow.

Setelah model sudah dibangun tahap terakhir yaitu dengan mengimplementasikannya ke dalam GUI (Graphical User Interface) menggunakan Streamlit

Pembahasan

Hasil pelatihan dan pengujian Model dapat dilihat pada Gambar 5 dan 6 dibawah. Model berhasil dilatih dengan akurasi sekitar 93.3% pada data latih dan mampu memberikan prediksi dengan akurasi sekitar 89.2% pada data uji. Hal ini menunjukkan bahwa model tidak mengalami overfitting secara signifikan dan dapat melakukan generalisasi pada data yang belum pernah dilihat sebelumnya dengan baik..

Tabel hasil klasifikasi memberikan gambaran tentang performa model dalam mengklasifikasikan berbagai kelas suara. Suara "air_conditioner" memiliki presisi sekitar 90% dengan recall 93%, Menunjukkan bahwa model cukup baik dalam mengenali suara tersebut. Kelas "car_horn" memiliki presisi yang sangat tinggi (99%) tetapi recall sedikit lebih rendah (88%), mungkin ada beberapa sampel yang tidak terdeteksi. Suara "children_playing" memiliki recall yang tinggi (90%), namun presisinya hanya sekitar 70%, yang menunjukkan ada kesulitan dalam membedakan suara ini dari kelas lain. Kelas "drilling" memiliki presisi sekitar 96% dengan recall 85%, menunjukkan model cenderung mengklasifikasikan beberapa suara drilling sebagai kelas lain. Sebagian besar kelas lain memiliki performa yang baik dengan presisi dan recall di atas 85%.

Pada Gambar 6, dapat dilihat hasil dari graph dengan melakukan percobaan sebanyak 100 epoch, hasil training dan validasi menunjukkan peningkatan performa model berkembang dengan baik selama 100 epoch berjalan dan penurunan training dan validation loss yang cukup konsisten.

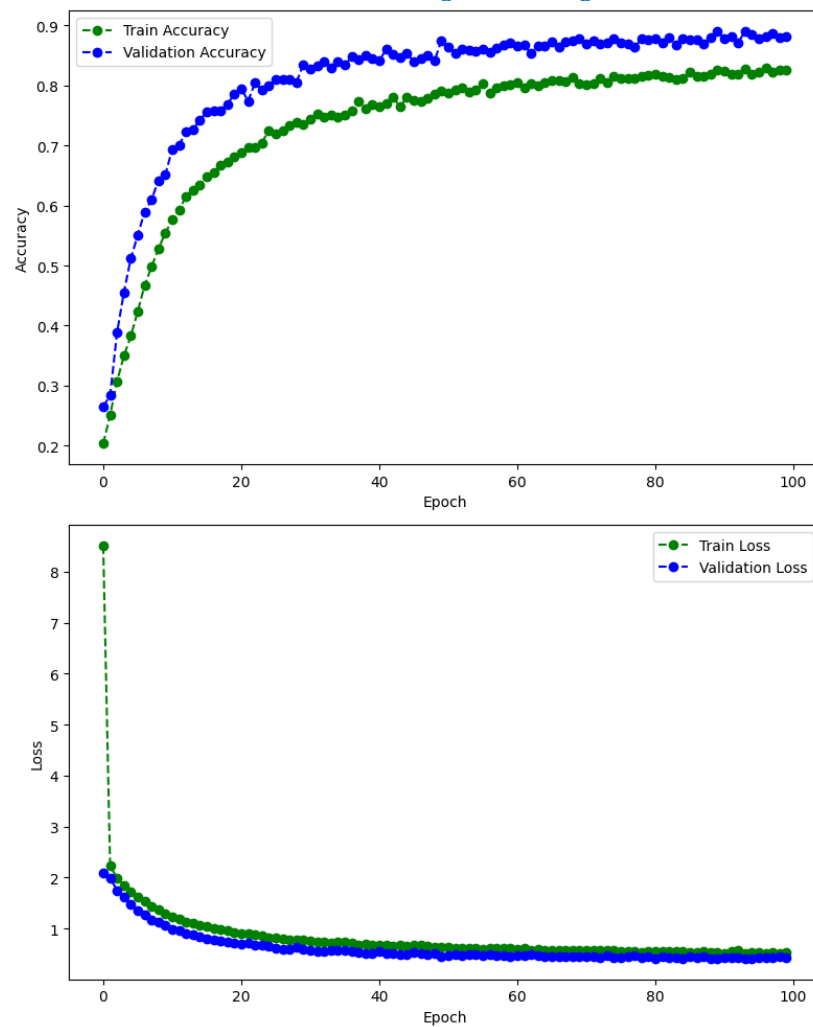
Pada Gambar 7 dapat dilihat dari hasil tabel confusion matriks prediksi, semua label sudah dapat memprediksi dengan baik namun masih terdapat beberapa kesalahan prediksi pada children_paling atau suara anak-anak yang sedang bermain sesuai dengan tingkat akurasi yang didapat pada gambar 5. Label tersebut memiliki tingkat akurasi yang rendah karena sulit menentukan suara anak-anak yang sedang bermain dengan kebisingan lingkungan yang lain.

Training Accuracy: 0.9255547523498535
 Testing Accuracy: 0.8809387683868408
 55/55 [=====] - 0s 1ms/step

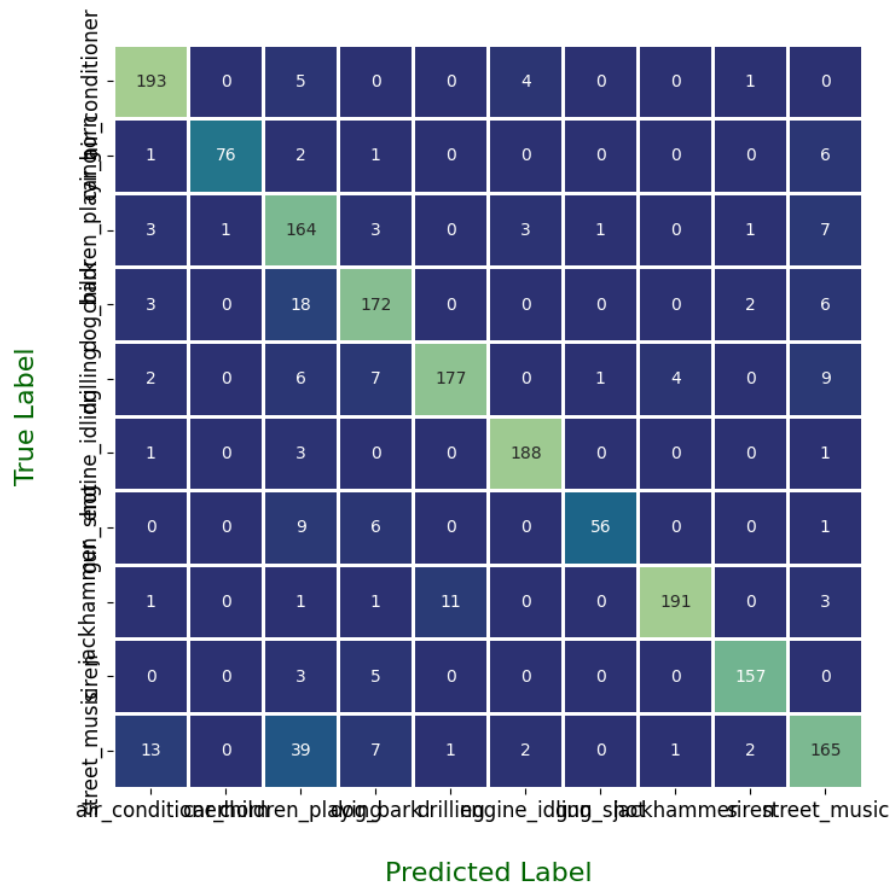
Classification Report :

	precision	recall	f1-score	support
air_conditioner	0.89	0.95	0.92	203
car_horn	0.99	0.88	0.93	86
children_playing	0.66	0.90	0.76	183
dog_bark	0.85	0.86	0.85	201
drilling	0.94	0.86	0.90	206
engine_idling	0.95	0.97	0.96	193
gun_shot	0.97	0.78	0.86	72
jackhammer	0.97	0.92	0.95	208
siren	0.96	0.95	0.96	165
street_music	0.83	0.72	0.77	230
accuracy			0.88	1747
macro avg	0.90	0.88	0.89	1747
weighted avg	0.89	0.88	0.88	1747

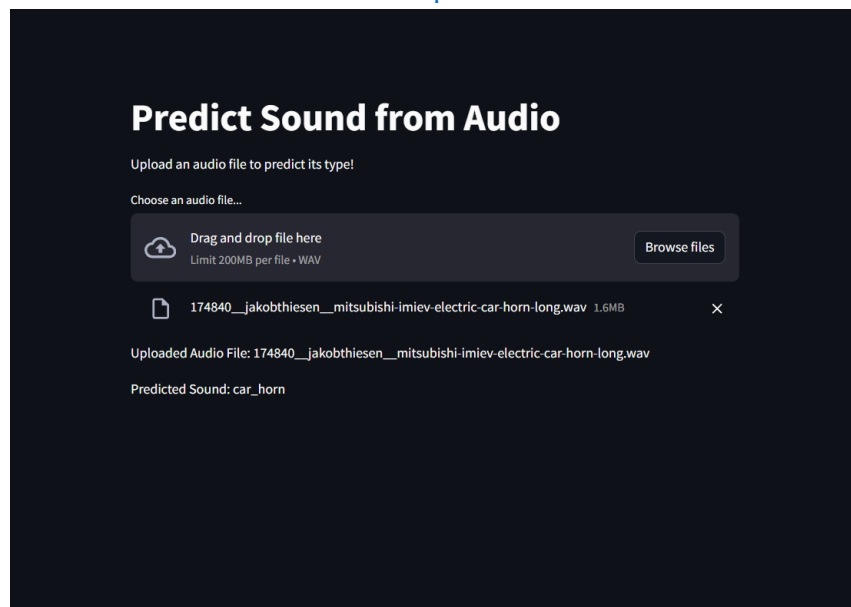
Gambar 5. Hasil Training dan Testing Model.



Gambar 6. Graph Pelatihan Model.



Gambar 7. Tabel Confusion Matriks Prediksi Model



Gambar 8. Hasil Implementasi GUI menggunakan Streamlit

4. SIMPULAN

Penelitian ini memanfaatkan MLP dan MFCC untuk mengklasifikasikan suara lingkungan perkotaan dengan tingkat akurasi signifikan. Melalui eksperimen pada dataset UrbanSound8K, metode ini berhasil mengidentifikasi sembilan jenis suara kota, meskipun masih ada beberapa tantangan dalam mengklasifikasikan beberapa kelas suara secara tepat. Model MLP yang digunakan menunjukkan kemampuan generalisasi yang baik dengan akurasi sekitar 93.3% pada data latih dan 89.2% pada data uji,

dan total akhir akurasi sebesar 89.2%. Penelitian ini memberikan kontribusi tambahan dalam bidang pemrosesan sinyal suara dan pembelajaran mendalam serta membuka peluang untuk pengembangan lebih lanjut. Saran untuk penelitian selanjutnya dapat mencakup eksplorasi menggunakan model deep learning alternatif seperti Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), dan pendekatan berbasis Transformers untuk klasifikasi suara perkotaan. Penelitian komparatif antara metode MLP yang digunakan dalam studi ini dengan model-model tersebut dapat memberikan wawasan lebih luas mengenai performa relatif dan keunggulan masing-masing model dalam mengklasifikasikan suara lingkungan perkotaan

5. DAFTAR PUSTAKA

- J. K. Das, A. Ghosh, A. K. Pal, S. Dutta, and A. Chakrabarty, "Urban Sound Classification Using Convolutional Neural Network and Long Short Term Memory Based on Multiple Features," *2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS)*, Oct. 2020, **Published**, doi: 10.1109/icds50568.2020.9268723.
- J. Salamon, C. Jacoby, and J. Bello, "Urban Sound Datasets.," <https://urbansounddataset.weebly.com/urbansound8k.htm>.
- Karthika, N., & Janet, B. (2020). Deep convolutional network for urbansound classification. *Sādhanā*, 45(1), 1-91
- Nogueira, A. F. R., Oliveira, H. S., Machado, J. J. M., & Tavares, J. M. R. S. (2022). Transformers for urban sound classification—a comprehensive performance evaluation. *Sensors*, 22(22), 88742
- S. Garg, T. Sehga, A. Jain, Y. Garg, P. Nagrath, and R. Jain, "RETRACTED: Urban Sound Classification Using Convolutional Neural Network Model," *IOP Conference Series: Materials Science and Engineering*, vol. 1099, no. 1, p. 012001, Mar. 2021, doi: 10.1088/1757-899x/1099/1/012001.
- S. Kilcommins, "Streamlit — Everything You Need To Know - DataDrivenInvestor," *Medium*, Jan. 01, 2022. <https://medium.datadriveninvestor.com/streamlit-everything-you-need-to-know-665eb90fcf4a#b992>