

Klasifikasi Audio Cats and Dogs Menggunakan Model *Artificial Neural Network Multi-perceptron*

Al Waasiu¹, Ahmad Ilham B², Armin Lawi³

^{1,2,3}Departemen Matematika, Universitas Hasanuddin

Email :¹waasiua18h@student.unhas.ac.id ²ahmadilhamb14@gmail.com ³lawi.armin@gmail.com

Abstract — *Deep learning is one part of various machine learning methods that use artificial neural networks (ANN). Deep Learning can detect types of animal sounds using ANN multi-perceptron. This paper presents a multi-perceptron ANN method for Classification of cat and dog sounds. The proposed ANN uses 3 hidden layers with the number of perceptrons 256, 128 and 64 neurons, respectively, and ends with 2 decisions whether the sound of a dog or a cat. The data used is 277 with 92 cat sounds and 185 dog sounds. The model uses 70% of the total data and 30% to test the performance of the model. The accuracy results are 88.04% with 83% precision and 86% sensitivity, where the positive label indicates the sound of a dog and the negative indicates a cat.*

Keyword — *Artificial Neural Network Multi-Perceptron, Classification, Data pre-processing, Deep Learning.*

Abstrak — *Deep Learning* merupakan salah satu bagian dari berbagai macam metode *Mechine Learning* yang menggunakan *Artificial Neural Networks* (ANN). *Deep Learning* dapat mendeteksi jenis-jenis suara binatang menggunakan ANN *multi-perceptron*. . Makalah ini menyajikan metode ANN *multi-perceptron* untuk Klasifikasi jenis suara kucing dan anjing. ANN yang diusulkan menggunakan 3 *hidden layer* dengan jumlah perceptron masing-masing 256, 128 dan 64 *neuron*, dan berakhir dengan 2 keputusan apakah suara anjing atau kucing. Data yang digunakan berjumlah 277 dengan suara kucing dan 185 suara anjing. Pembentukan model menggunakan 70% dari jumlah data dan 30% untuk pengujian mengukur kinerja model. Adapun hasil akurasi adalah sebesar 88,04% dengan presisi 83% dan sensitifitas 86%, dimana label positif menyatakan suara anjing dan negatif menyatakan kucing.

Kata kunci — *Artificial Neural Networks multi-perceptron, Deep Learning, Data pre-processing, Klasifikasi.*

I. PENDAHULUAN

Berdasarkan MIT Technology Review, Deep Learning memiliki artian yaitu pembelajaran yang mendalam dan juga dikenal sebagai pembelajaran terstruktur dalam atau pembelajaran hierarkis. Dengan arti lain, Deep Learning merupakan sebuah model yang dapat mempelajari metode komputasinya sendiri menggunakan ‘otaknya’ sendiri. Algoritma Deep Learning menggunakan metadata sebagai input dan mengolahnya menggunakan sejumlah lapisan tersembunyi (hidden layer) transformasi non linier dari data masukan untuk menghitung nilai output. Algoritma pada Deep Learning memiliki fitur yang unik yaitu sebuah fitur

yang mampu mengekstraksi secara otomatis. Artificial Neural Networks yang biasa disingkat dengan ANN adalah bagian yang paling peting dari deep learning. ANN ini mensimulasikan kerja otak kita yang tersusun jaringan saraf yang disebut neuron. Sama seperti sistem otak manusia, dalam jaringan artificial neural network ini si mesin menerima informasi pada titik-titik yang disebut nodes yang terkumpul pada satu layer untuk kemudian diteruskan dan diproses ke layer selanjutnya yang disebut hidden layers. Dataset yang digunakan adalah Audio Cats and Dogs berisi tentang Dataset terdiri dari banyak file "wav" untuk kelas cats dan dogs:

- cats memiliki 164 file WAV yang sesuai dengan audio 1323 detik.
- dogs memiliki 113 file WAV yang sesuai dengan audio 598 detik .

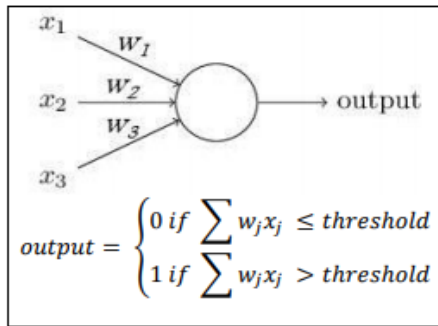
II. TINJAUAN PUSTAKA

A. Landasan Teori atau Konsep Dasar

Deep learning merupakan salah satu bagian dari berbagai macam metode machine learning yang menggunakan artificial neural networks (ANN).

1. Definisi Artificial Neural Networks (ANN)
Artificial Neural Network (ANN) atau jaringan syaraf tiruan adalah jaringan dari sekelompok unit pemroses kecil yang dimodelkan berdasarkan perilaku jaringan syaraf manusia (Wikipedia).

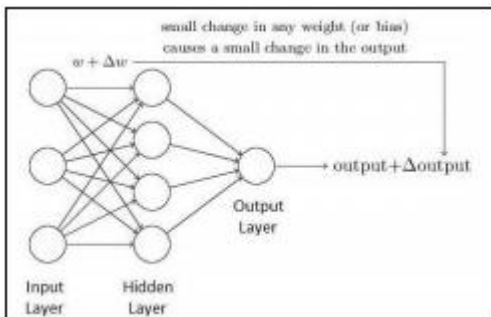
2. Artificial Neural Network *Multi-Perceptron*
Perceptron pada dasarnya adalah gerbang logika, di mana banyak input diambil, dibobot secara individual, dan dijumlahkan. Jika jumlah input berbobot lebih besar dari nilai ambang batas, maka output dari perceptron adalah “1”, sebaliknya outputnya adalah “0”. Hal ini ditunjukkan secara grafis pada Gambar 1 (grafik diadaptasi dari Nielson).



Gambar 1. Logika Perceptron

Perceptron kemudian disempurnakan untuk memungkinkan keluaran nonbiner (yaitu keluaran selain hanya 1 dan 0). Hal ini dicapai dengan mengganti logika fungsi langkah dengan fungsi sigmoid. Untuk alasan ini, model baru untuk perceptron diberi nama "neuron sigmoid". Bagian "neuron" dari nama tersebut berasal dari kesamaan neuron sigmoid dalam operasinya dengan neuron biologis, yang pada tingkat yang sangat dasar mengaktifkan output dengan input yang cukup.

Neuron sigmoid menjadi berguna ketika dihubungkan bersama untuk melakukan logika kompleks, dan dinamai "neuron tersembunyi" dalam lingkup jaringan saraf. Gambar 2 (grafik diadaptasi dari Nielson) menunjukkan ANN yang berisi 3 input, 4 neuron tersembunyi, dan satu output.



Gambar 2. ANN sederhana

Dalam persamaan fungsi sigmoid dari fungsi logistik didefinisikan oleh rumus:

$$g(z) = \frac{1}{1 + e^{-z}}$$

Gambar 3. Rumus fungsi sigmoid

B. Penelitian Terdahulu

- 1) Deep learning beyond cats and dogs: recent advances in diagnosing breast cancer with deep neural networks

Deep learning has demonstrated tremendous revolutionary changes in the computing industry and its effects in radiology and imaging sciences have begun to dramatically change screening paradigms. Specifically, these advances have influenced the development of computer-aided detection and diagnosis (CAD) systems. These technologies have long been thought of as "second-opinion" tools for radiologists and clinicians. However, with significant improvements in deep neural networks, the diagnostic capabilities of learning algorithms are

approaching levels of human expertise (radiologists, clinicians etc.), shifting the CAD paradigm from a "second opinion" tool to a more collaborative utility. This paper reviews recently developed CAD systems based on deep learning technologies for breast cancer diagnosis, explains their superiorities with respect to previously established systems, defines the methodologies behind the improved achievements including algorithmic developments, and describes remaining challenges in breast cancer screening and diagnosis. We also discuss possible future directions for new CAD models that continue to change as artificial intelligence algorithms evolve.

- 2) Evaluating deep learning architectures for Speech Emotion Recognition

Speech Emotion Recognition (SER) can be regarded as a static or dynamic classification problem, which makes SER an excellent test bed for investigating and comparing various deep learning architectures. We describe a frame-based formulation to SER that relies on minimal speech processing and end-to-end deep learning to model intra-utterance dynamics. We use the proposed SER system to empirically explore feed-forward and recurrent neural network architectures and their variants. Experiments conducted illuminate the advantages and limitations of these architectures in paralinguistic speech recognition and emotion recognition in particular. As a result of our exploration, we report state-of-the-art results on the IEMOCAP database for speaker-independent SER and present quantitative and qualitative assessments of the models' performances.

- 3) Performance Analysis of Malayalam Language Speech Emotion Recognition System Using ANN/SVM

Automatic recognition of emotions from speech by machines has been one of the most challenging areas of research in the field of human machine interaction. Automatic emotion recognition system by speech merely means that to monitor and identify the emotional or physiological state of an individual from their utterances. Speech emotion recognition has wide range of application ranging from clinical studies to robotics. In this paper developed speech emotional database for Malayalam language (One of the south Indian languages) and a system for recognizing the emotions. The system used Mel Frequency Cepstral Coefficients (MFCCs), Short Time Energy (STE) and Pitch as features extraction techniques. Two classifiers, namely Artificial Neural Network (ANN) and Support Vector Machine (SVM) used for pattern classification. Experiments show that this method provides a high accuracy of 88.4% in the case of ANN and 78.2% in the case of SVM.

III. METODE PENELITIAN

A. Penjelasan Metode

1. Analisis eksplorasi data
Menginput data dan menganalisis data dengan bantuan visualisasi dan perhitungan statistik pada data.
2. Praproses Data

- Categorical encoding dengan label encoding dan one hot encoding

Label-Encoding mengacu pada perubahan label menjadi bentuk numerik untuk mengubahnya menjadi bentuk yang dapat dibaca mesin (*computer*). *Label-Encoding* mengonversi data dalam bentuk yang dapat dibaca mesin, tetapi memberikan nomor unik (mulai dari 0) untuk setiap kelas data. *One-Hot-Encoding* adalah proses di mana variabel kategori diubah menjadi kolom-kolom baru, Setiap kolom berisi "0" atau "1" yang sesuai dengan kolom mana label atau kategori tersebut berada.

3. Reduksi dimensi

- Feature Extraction

Secara sederhana, tujuan dari PCA adalah mengurangi jumlah atribut pada dataset tanpa mengurangi informasi. Contohnya pada sebuah dataset harga rumah. Pada PCA setiap atribut disebut sebagai principal component. Jika terdapat 10 atribut pada dataset, berarti terdapat 10 principal component. Pada gambar di bawah terdapat histogram dari 10 principal component dan variance dari setiap principal component

4. Proses Klasifikasi

- Artificial Neural Network (ANN)

Artificial Neural Network (ANN) atau jaringan syaraf tiruan adalah jaringan dari sekelompok unit pemroses kecil yang dimodelkan berdasarkan perilaku jaringan syaraf manusia (Wikipedia). Algoritma ANN lahir dari gagasan seorang psikolog Warren McCulloch dan Walter Pitts pada 1943 yang menjelaskan cara kerja jaringan syaraf dengan perangkat jaringan elektronik. Dapat di lihat pada Gambar 2.

5. Pengukuran Performance

- Confusion Matrix

Confusion Matrix adalah pengukuran performa untuk masalah klasifikasi machine learning dimana keluaran dapat berupa dua kelas atau lebih. Confusion Matrix adalah tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai aktual. Ada empat istilah yang merupakan representasi hasil proses klasifikasi pada confusion matrix yaitu True Positif, True Negatif, False Positif, dan False Negatif.

- Accuracy

Merupakan rasio prediksi Benar (positif dan negatif) dengan keseluruhan data

$$\text{Akurasi} = (TP + TN) / (TP+FP+FN+TN)$$

- Precision

Merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif.

$$\text{Precision} = (TP) / (TP+FP)$$

- Recall (Sensitifitas)

Merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif.

$$\text{Recall} = (TP) / (TP + FN)$$

- Specificity

Merupakan kebenaran memprediksi negatif dibandingkan dengan keseluruhan data negatif.

$$\text{Specificity} = (TN) / (TN + FP)$$

- F1 Score

F1 Score merupakan perbandingan rata-rata presisi dan recall yang dibobotkan

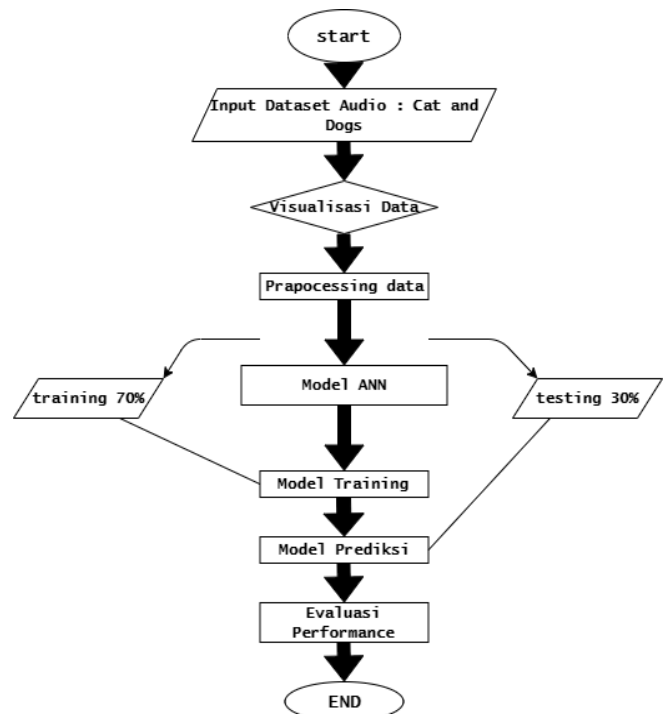
$$F1 \text{ Score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$$

B. Lokasi dan Waktu

Waktu Pengerjaan	Aktifitas Yang Di Lakukan	Lokasi
20/05/2021	Mencari dataset di kaggle	Rumah Alamat : Raha, Sulawesi Tenggara.
21/05/2021	Import dataset di Google Colab	Rumah Alamat : Raha, Sulawesi Tenggara.
22/05/2021	Prapocessing dataset dan buat model ANN di Google Colab	Rumah Alamat : Raha, Sulawesi Tenggara.
29/05/2021	Selesai	Rumah Alamat : Raha, Sulawesi Tenggara.

Gambar 4. Lokasi dan Waktu pengerjaan

C. Alur Kerja



Gambar 5. Alur Kerja

D. Sumber Data

Pada penelitian ini, kami mengambil dataset dari repository kaggle. Dataset yang kami gunakan pada penelitian ini adalah dataset Audio Cats and Dogs berisi tentang Dataset terdiri dari banyak file "wav" untuk kelas cats dan dogs:

Jenis sound	Tipe sound	Ukuran sound
Cats	164	1323 detik
Dogs	113	598 detik

E. Instrumen Penelitian

– Perangkat Keras

Jenis perangkat yang digunakan untuk mengolah dataset adalah laptop Acer Intel CORE i3 dengan spesifikasi sebagai berikut:

- Processor : Intel(R) Core(TM) i3-5005U @ 2.00GHz
- RAM : 10 GB DDR3L 1333
- Operating System : Windows 10 Home Single Language 64-bit (10.0, Build 19041)

– Perangkat Lunak

Perangkat lunak yang kami gunakan untuk menguji dataset Audio Cats and Dogs dari model Artificial Neural Network (ANN) yaitu Google Colab.

IV. HASIL DAN PEMBAHASAN

Total ada 277 sampel yang terbagi dalam 2 kelas :

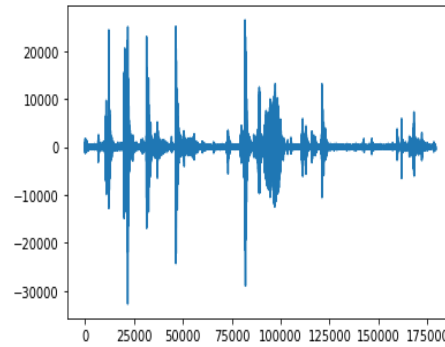
- Cat memiliki 164 file WAV yang sesuai dengan audio 1323 detik.
 - Anjing memiliki 113 file WAV yang sesuai dengan audio 598 detik.
 - Dari google drive import dataset audio ke google colab
 - Import package
 - Lokasi penyimpanan folder terletak didalam ("[/content/drive/MyDrive/dataset/audio-cats-and-dogs/cats_dogs/](#)")
berisi file berikut : ['utils.py', 'train_test_split.csv', 'cats_dogs']
 - Praprocessing dataset
- isi keseluruhan dataset audio cats and dogs = 277
- Spliting Data Train dan Data Test
Sebelum masuk ke bagian klasifikasi terlebih dahulu kita menentukan pembagian jumlah data training dan data testing yang akan di gunakan.

```
# Split train and test
X_train, X_test, y_train, y_test = train_test_split(X_path, y,
test_size=0.33)

print("in X, there is {} cats and {} dogs".format(len(y) -
sum(y), sum(y)))
print("in X_train, there is {} cats and {} dogs".format(len(y_
_train) - sum(y_train), sum(y_train)))
print("in X_test, there is {} cats and {} dogs".format(len(y_t
est) - sum(y_test), sum(y_test)))
print("len(y_train)",len(y_train))
print("len(y_test)",len(y_test))
print(y_test.shape)
print(y_train.shape)
```

```
in X, there is [164.] cats and [113.] dogs
in X_train, there is [110.] cats and [75.] dogs
in X_test, there is [54.] cats and [38.] dogs
len(y_train) 185
len(y_test) 92
(92, 1)
(185, 1)
```

➤ Melihat sinyal data audio

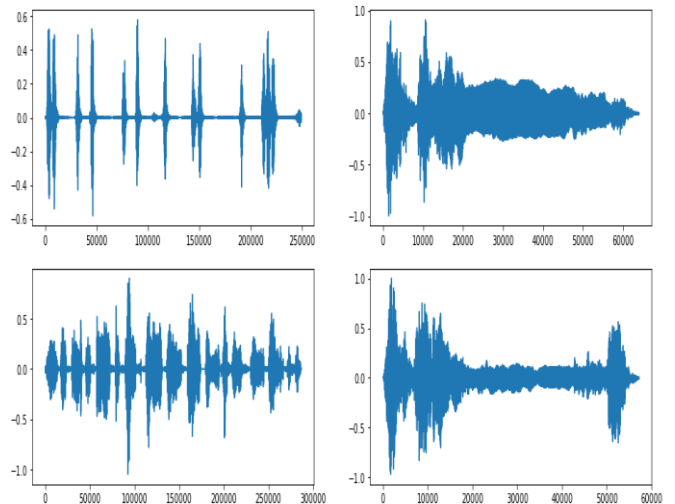


cat_1.wav
(16000, array([124, 107, 7, ..., -545, -494, -529], dtype=int16))

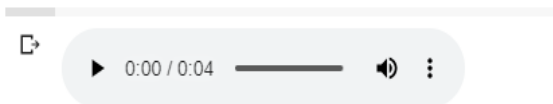
➤ Menampilkan Plot sinyal _train audio .wav

```
import matplotlib.pyplot as plt

fig, axs = plt.subplots(2, 2, figsize=(16,7))
axs[0][0].plot(X_train[0])
axs[0][1].plot(X_train[1])
axs[1][0].plot(X_train[2])
axs[1][1].plot(X_train[3])
plt.show()
```

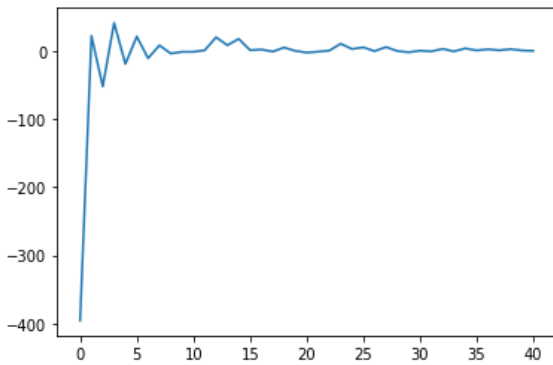


➤ menampilkan Suara dari dog_barking_15.wav



➤ Megestrak semua fitur untuk di klasifikasi Gambar plot klasifikasi

```
1
(1, 41)
[<matplotlib.lines.Line2D at 0x7f8a68216e10>]
```



```
scores = model.evaluate(X_test_features, test_labels)
print(f"Test Accuracy: {scores[1]*100}")
```

```
3/3 [=====] - 0s 3ms/step - loss: 0.4811 - accuracy: 0.8804
Test Accuracy: 88.04348111152649
```

- Simpan di 'model_keras.h5'
- List semua data history dengan :
 - label = "training accuracy"
 - label = "validation accuracy"

&

```
dict_keys(['loss', 'accuracy', 'val_loss', 'val_accuracy'])
```

- Daftar array shape dari ekstrak fitur

Image array shape: (185, 41)

Image array shape: (92, 41)

Label array shape: (185, 1)

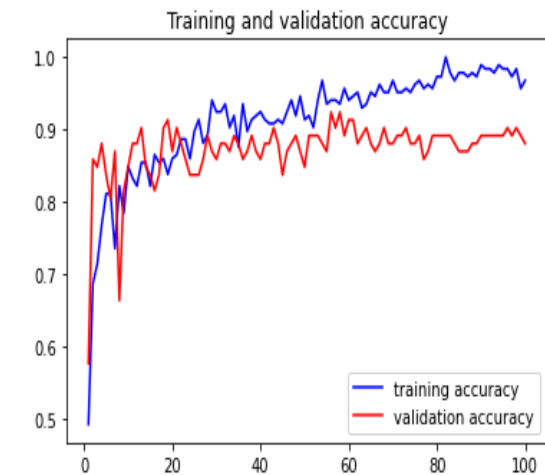
Label array shape: (92, 1)

- Mengkonversi label ke to_categorical
- Membuat model Artificial Neural Network menggunakan 3 hidden layer dengan jumlah perceptron masing-masing 256, 128 dan 64 neuron dengan menerapkan DropOut

```
Model: "sequential"
```

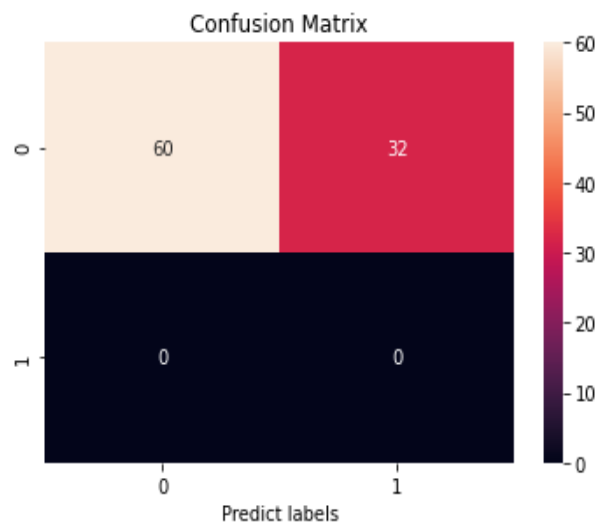
Layer (type)	Output Shape	Param #
dense (Dense)	(None, 256)	10752
dense_1 (Dense)	(None, 128)	32896
dropout (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 64)	8256
dropout_1 (Dropout)	(None, 64)	0
dense_3 (Dense)	(None, 2)	130

```
Total params: 52,034
Trainable params: 52,034
Non-trainable params: 0
```



- Gambar Confusion Marix dari klasifikasi menggunakan metode ANN Back-Prpagation Multi-Perceptron

```
Text(0.5, 1.0, 'Confusion Matrix')
```



- Hasil akuransi di epoch =100 adalah 88,04%

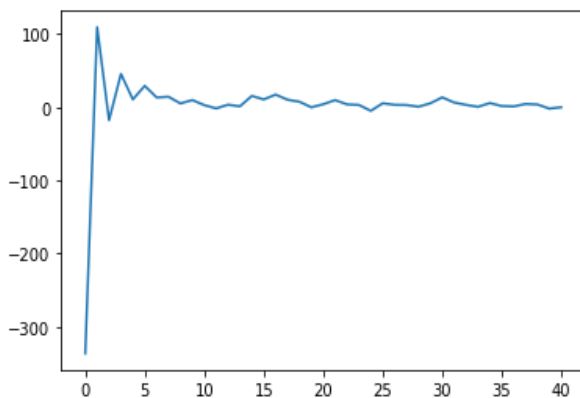
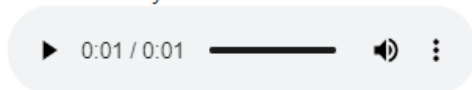
```
WARNING:tensorflow:Can save best model only with val_acc available, skipping.
Epoch 92/100
6/6 [=====] - 0s 6ms/step - loss: 0.0721 - accuracy: 0.9786 - val_loss: 0.4587 - val_accuracy: 0.8804
WARNING:tensorflow:Can save best model only with val_acc available, skipping.
Epoch 93/100
6/6 [=====] - 0s 10ms/step - loss: 0.0277 - accuracy: 0.9868 - val_loss: 0.4321 - val_accuracy: 0.8804
WARNING:tensorflow:Can save best model only with val_acc available, skipping.
Epoch 94/100
6/6 [=====] - 0s 6ms/step - loss: 0.0473 - accuracy: 0.9912 - val_loss: 0.4354 - val_accuracy: 0.8804
WARNING:tensorflow:Can save best model only with val_acc available, skipping.
Epoch 95/100
6/6 [=====] - 0s 8ms/step - loss: 0.0474 - accuracy: 0.9832 - val_loss: 0.4297 - val_accuracy: 0.8804
WARNING:tensorflow:Can save best model only with val_acc available, skipping.
Epoch 96/100
6/6 [=====] - 0s 7ms/step - loss: 0.0682 - accuracy: 0.9817 - val_loss: 0.4640 - val_accuracy: 0.8804
WARNING:tensorflow:Can save best model only with val_acc available, skipping.
Epoch 97/100
6/6 [=====] - 0s 6ms/step - loss: 0.0710 - accuracy: 0.9685 - val_loss: 0.4640 - val_accuracy: 0.8804
WARNING:tensorflow:Can save best model only with val_acc available, skipping.
Epoch 98/100
6/6 [=====] - 0s 6ms/step - loss: 0.0362 - accuracy: 0.9896 - val_loss: 0.4792 - val_accuracy: 0.8804
WARNING:tensorflow:Can save best model only with val_acc available, skipping.
Epoch 99/100
6/6 [=====] - 0s 7ms/step - loss: 0.0579 - accuracy: 0.9622 - val_loss: 0.4645 - val_accuracy: 0.8804
WARNING:tensorflow:Can save best model only with val_acc available, skipping.
Epoch 100/100
6/6 [=====] - 0s 7ms/step - loss: 0.0692 - accuracy: 0.9784 - val_loss: 0.4811 - val_accuracy: 0.8804
WARNING:tensorflow:Can save best model only with val_acc available, skipping.
```

- Hasil Precision = 83%, Sensitive/recall = 86%, F-1 Score = 84%

	precision	recall	f1-score	support
0	0.89	0.87	0.88	39
1	0.83	0.86	0.84	28
accuracy			0.87	67
macro avg	0.86	0.86	0.86	67
weighted avg	0.87	0.87	0.87	67

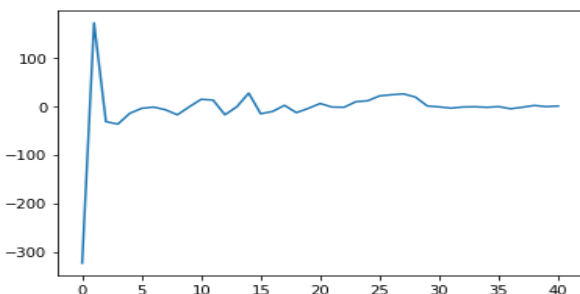
➤ Hasil prediksi = 1 didapatkan Cat: 0.9996252059936523 Dog: 0.00037474147393368185 karena cats lebih besar dari dog sehingga sinyal memberikan suara cats

Cat: 0.9996252059936523 Dog: 0.00037474147393368185
The label says that it is a Cat!



➤ Hasil prediksi = 50 didapatkan Dog: 1.1572664604386773e-08 Cat: 1.0 karena cats lebih kecil dari dog sehingga sinyal memberikan suara dogs

Dog: 1.1572664604386773e-08 Cat: 1.0
The label says that it is a Dog!



V. KESIMPULAN

Penelitian yang dilakukan menggunakan dataset kucing dan anjing dengan ukuran file sebesar 97 MB, terdapat 2 kelas jenis file "wav" antara lain : Cats, and Dogs.

Langkah-langkah yang dilakukan pada penelitian adalah preprocessing data, yaitu memproses data agar sesuai dengan kriteria pada model pada preprocessing data

dilakukan tiga proses yaitu mengacak urutan set pelatihan, memisahkan data training dan data testing dengan perbandingan 70:30, Mengubah audio menyesuaikan model ANN. Setelah preprocessing data, dilakukan proses klasifikasi, pada proses ini dilakukan training dan testing pada model, prediksi model terhadap data baru. Kemudian proses selanjutnya melakukan evaluasi kinerja dengan metode Accuracy, Precision, Recall, F-Score, dan *confusion matrix*. preprocessing data, dilakukan proses klasifikasi, pada proses ini dilakukan training dan testing pada model, prediksi model terhadap data baru. Kemudian proses selanjutnya melakukan evaluasi kinerja dengan metode Accuracy, Precision, Recall, F-Score, dan *confusion matrix*.

VI. DAFTAR PUSTAKA

- [1] Audio Cats and Dogs. Classify raw sound events: <https://www.kaggle.com/mmoreaux/audio-cats-and-dogs>
- [2] Mengenal Deep Learning Untuk Yang Tak Mau Pusing Aug 21, 2019 : <https://inixindojogja.co.id/mengenal-deep-learning/>
- [3] Neural Networks With Backpropagation For Xor Using One Hidden Layer: https://www.bogotobogo.com/python/python_Neural_Networks_Backpropagation_for_XOR_using_one_hidden_layer.php
- [4] Statement A: Approved for Release. Distribution is unlimited: <https://arxiv.org/ftp/arxiv/papers/1603/1603.05189.pdf>