

Klasifikasi Suara Kucing dan Anjing Menggunakan LSTM-GRU dan ANN-BP

Fuad Hamdi Bahar¹, Nasmah Indah Sari² dan Armin Lawi³

^{1,2,3} Sistem Informasi, Universitas Hasanuddin

fuadhamdi99@gmail.com¹, nasmahindah.sari@gmail.com², armin@unhas.ac.id³

Abstract — Cats and dogs are animals that many people keep. The amount of data used is 277 data, 164 cat sounds and 113 dog sounds, which are then divided into 80% training data and 20% testing data. To distinguish the voices of the two animals, the LSTM-GRU and ANN-BP methods were used. The goal to be achieved is to carry out machine learning by proposing the sound of a cat or dog and then we ask the machine whether it is a cat or dog sound. The machine's thought process is based on the algorithm we created. The accuracy results are 92% with a precision of 0.91 and a recall of 0.91.

Keyword — Deep Learning, Long Short-Term Memory, Gated Recurrent Unit, Artificial Neural Network, Backpropagation.

Abstrak — Kucing dan anjing merupakan binatang yang banyak dipelihara masyarakat. Jumlah data yang digunakan sebanyak 277 data, suara kucing sebanyak 164 dan suara anjing 113 data yang kemudian dibagi kedalam 80% data training dan 20% data testing. Untuk membedakan suara kedua binatang tersebut digunakan metode LSTM-GRU dan ANN-BP. Tujuan yang ingin dicapai ialah melakukan pembelajaran yang dilakukan oleh mesin dengan cara menyodorkan suara seekor kucing atau anjing lalu kita bertanya pada mesin tersebut apakah itu suara kucing atau anjing. Proses berpikir mesin tersebut didasari oleh algoritma yang kita buat. Adapun hasil akurasi adalah sebesar 92% dengan precision 0.91 dan recall 0.91.

Kata kunci — Deep Learning, Long Short-Term Memory, Gated Recurrent Unit, Artificial Neural Network, Backpropagation.

pada klasifikasi suara hewan dan identifikasi populasi hewan berdasarkan karakteristik suaranya. Klasifikasi suara mamalia laut dan dampaknya terhadap kehidupan laut. Klasifikasi spesies serangga berdasarkan sinyal suara mereka menggunakan deep learning. Prediksi perilaku suara hewan yang tidak biasa selama gempa bumi dan bencana alam dipelajari menggunakan deep learning [2].

Anjing dan Kucing adalah 2 jenis hewan yang banyak dipelihara manusia sebagai peliharaan rumah. Kedua jenis hewan ini berkomunikasi menggunakan suara. Cukup mudah untuk membedakan suara anjing dan juga kucing secara langsung. Namun dirasa cukup menarik untuk melakukan pengklasifikasian suara anjing dan juga kucing menggunakan metode deep learning.

Dalam penelitian ini, klasifikasi suara anjing dan kucing dilakukan dengan menggunakan model LSTM-GRU dan ANN-BP. LSTM-GRU sendiri merupakan salah satu metode dari Recurrent Neural Network yang biasanya digunakan dalam pengenalan suara, named entity recognition dan juga termasuk dalam hal klasifikasi. Sedangkan ANN adalah model machine learning yang multiguna, ANN dapat dipakai dalam klasifikasi gambar, mengenali ratusan Bahasa dunia, rekomendasi video dan lain-lain. Sedangkan Backpropagation adalah sebuah algoritma yang dapat membuat prediksi dari model menjadi semakin baik/akurat.

I. PENDAHULUAN

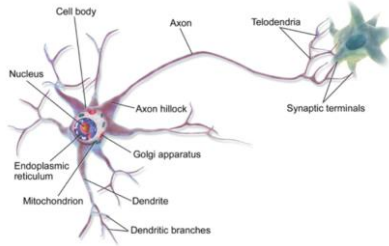
Selama dekade terakhir, penelitian dalam klasifikasi dan pengenalan suara semakin populer dan dengan cepat memperluas penerapannya dari fokus yang lebih tradisional pada pengenalan suara dan klasifikasi genre musik untuk identifikasi biometrik, deteksi suara jantung berbantuan computer, klasifikasi aksen bahasa Inggris, dan identifikasi gender. Seperti halnya penelitian pengenalan pola pada umumnya, fitur yang dimasukkan ke dalam pengklasifikasi pada awalnya direkayasa, yang dalam hal aplikasi suara, berarti mengekstrak dari jejak audio mentah deskriptor seperti Statistical Spectrum Descriptor dan Histogram Irama [1].

Suara hewan peliharaan apa pun bisa sangat membantu manusia dalam perspektif keamanan atau prediksi dari bencana alam. Klasifikasi dan pengenalan suara hewan peliharaan menggunakan deep learning dapat menjadi area penelitian yang berkembang. Pendekatan berbasis data untuk akustik sinyal mendapat minat yang tinggi dalam beberapa tahun terakhir dan beberapa penelitian juga fokus

II. LANDASAN TEORI

A. Artificial Neural Network

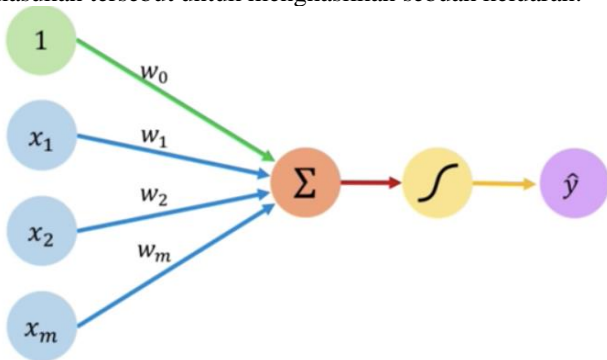
Jaringan Saraf Tiruan atau Artificial Neural Network (ANN) adalah sebuah model machine learning yang terinspirasi dari neuron/saraf yang terdapat pada otak manusia. Sebuah saraf terdiri dari 3 (tiga) bagian utama, yaitu badan sel yang disebut nukleus, akson, dan dendrit. Nukleus berisi materi genetik dan bertugas mengontrol seluruh aktivitas sel. Akson adalah cabang yang terlihat seperti ekor yang panjang, bertugas mengirimkan pesan dari sel. Sedangkan dendrit adalah cabang-cabang pendek yang terlihat seperti cabang pohon, bertugas menerima pesan untuk sel.



Setiap ujung akson dari sebuah neuron terhubung dengan dendrit dari neuron lainnya. Neuron berkomunikasi satu sama lain dengan mengirimkan senyawa kimia yang disebut neurotransmitter, melintasi ruang kecil (synapse) antara akson dan dendrit neuron yang berdekatan. Ketika sebuah neuron mendapatkan rangsangan, neuron tersebut akan mengirim sinyal ke neuron lainnya.

Perceptron adalah komponen dasar pembangun jaringan saraf tiruan. Perceptron pada jaringan saraf tiruan terinspirasi dari neuron pada jaringan saraf di otak manusia. Pada jaringan saraf tiruan, perceptron dan neuron merujuk pada hal yang sama.

Sebuah perceptron terdiri dari 5 komponen yaitu Input (x_i), Bobot atau weights (W_i) dan bias (W_0), Penjumlahan atau sum (Σ), Fungsi aktivasi atau non linearity function (f) dan Output (y). Sebuah perceptron menerima masukan berupa bilangan numerik. Perceptron kemudian memproses masukan tersebut untuk menghasilkan sebuah keluaran.



Inputs Weights Sum Non-Linearity Output
Fungsi matematis dari perceptron dapat dilihat dari gambar dibawah:

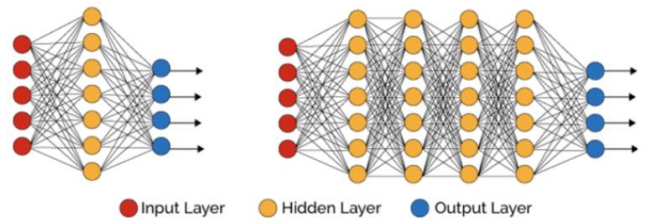
$$\hat{y} = g \left(w_0 + \sum_{i=1}^m x_i w_i \right)$$

Linear combination of inputs
Bias

Non-linear activation function

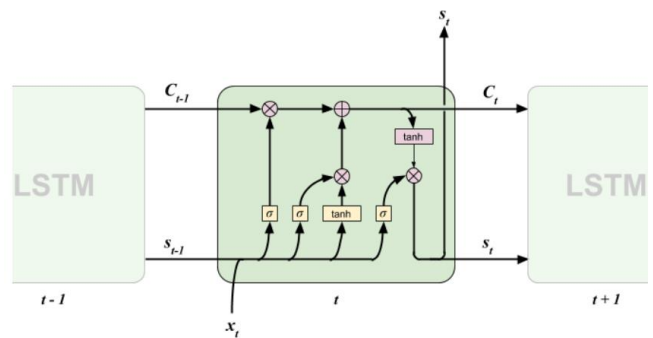
Secara umum, sebuah ANN terdiri atas Input layer, Hidden layer dan Output layer. Dalam sebuah jaringan saraf tiruan, input layer dan output layer harus selalu ada, namun untuk hidden layer bisa ada beberapa atau tidak sama sekali. Hidden layer dan output layer sama-sama merupakan sebuah layer yang memiliki beberapa perceptron. Sedangkan input layer adalah sebuah layer yang hanya menampung angka-angka.

Hidden layer diberikan nama hidden karena sifatnya yang tersembunyi. Pada sebuah sistem jaringan saraf, input dan output layer merupakan lapisan yang dapat kita amati, sementara hidden layer, tidak.

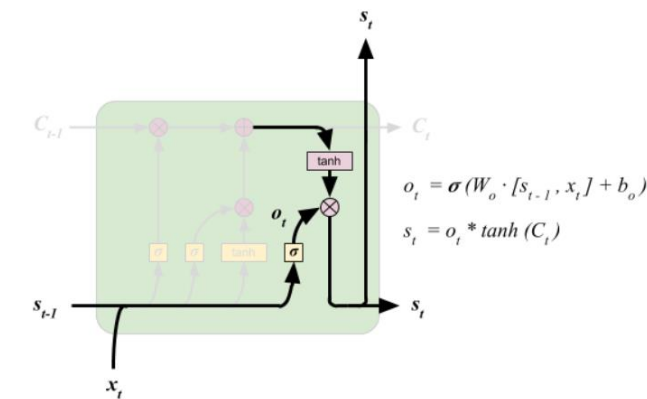
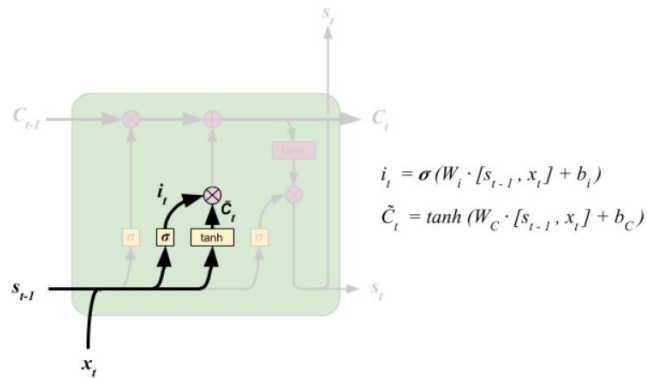
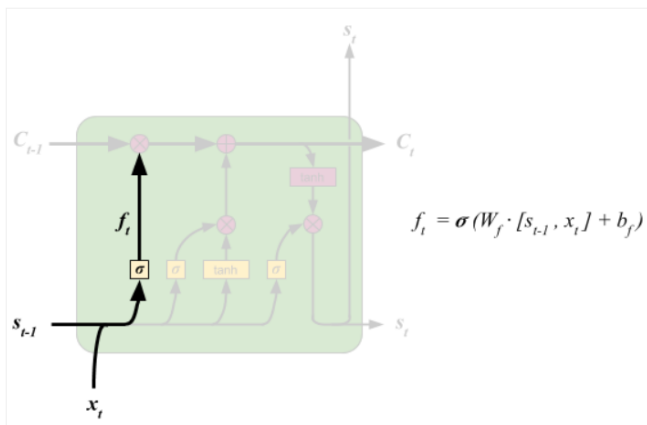


B. Long Short Term Memory

LSTM (Long Short Term Memory) merupakan jenis lain yang dari Recurrent Neural Network. Karena LSTM juga merupakan RNN maka didalamnya juga terdapat pemrosesan yang berulang. Gambar berikut menunjukkan arsitektur dari LSTM:

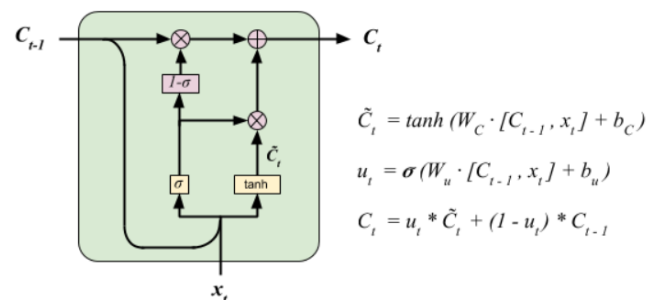


LSTM terdiri atas 3 gerbang atau gate, yaitu input gate, forget gate dan output gate. Input gate memutuskan nilai mana yang akan kita perbarui. Output gate memutuskan informasi apa yang akan dibuang dari konteks. Sedangkan Output gate memutuskan bagian-bagian apa dari konteks yang akan kita hasilkan. *Gated Recurrent Unit.*



C. Gated Recurrent Unit

Gated Recurrent Unit merupakan salah satu varian yang populer dari sekian banyaknya varian LSTM. Keutamaan GRU adalah komputasinya lebih sederhana dari LSTM, namun mempunyai akurasi yang setara dan masih cukup efektif untuk menghindari permasalahan gradien yang



menghilang. Gambar berikut menunjukkan arsitektur GRU:

Jika LSTM memiliki 3 gate maka GRU hanya memiliki 2 gate yaitu reset gate dan update gate. Update gate membantu model untuk menentukan berapa banyak informasi masa lalu (dari langkah waktu sebelumnya) yang perlu diteruskan ke masa depan. Sedangkan reset gate digunakan model untuk memutuskan berapa banyak informasi masa lalu untuk dilupakan.

III. METODE PENELITIAN

Lokasi Penelitian dilaksanakan di Kota Makassar pada tanggal 22 Mei 2021. Penelitian ini dilakukan untuk melakukan klasifikasi suara kucing dan anjing dengan menggunakan LSTM GRU pada arsitektur ANN-BP.

Data yang digunakan dalam penelitian ini adalah data yang diambil dari Repository Kaggle: <https://www.kaggle.com/mmoreaux/audio-cats-and-dogs>.

Data yang digunakan merupakan data suara kucing dan anjing (audio cats and dogs) yang diupload oleh Marc Moreaux. Dataset ini terdiri dari 277 file dalam format(.wav). Dalam dataset ini ada 2 kelas dengan rincian dalam bentuk table sebagai berikut:

No	Label	Jumlah Suara
1.	Cats	164
2.	Dogs	113

Perangkat lunak yang kami gunakan untuk menguji dataset Audio Cats and Dogs dari model Artificial Neural Network (ANN) yaitu Google Colab. Adapun perangkat keras yang digunakan dalam penelitian ini yaitu Laptop dengan spesifikasi processor Intel® Core™ i5-9300H CPU @2.40GHz, RAM : 8.00 GB dan system operasi Windows 10 Home Single Language 64-bit.

Indeks validasi berisi nilai-nilai yang menggambarkan keberhasilan atau kegagalan suatu proses. Proses yang dinilai dengan indeks ini yaitu proses klasifikasi nilai-nilai sinyal suara hasil ekstraksi fitur. Dengan melakukan validasi hasil klasifikasi, maka akan menunjukkan keberhasilan proses dengan pertimbangan berikut:

A. Confusion Matrix

Confusion matrix adalah matriks yang memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh model dengan hasil klasifikasi sebenarnya. Representasi hasil proses klasifikasi pada confusion matrix terbagi dalam 4 bagian, yaitu True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN).

		Nilai Aktual	
		Positive	Negative
Nilai Prediksi	Positive	TP (True Positive)	FP (False Positive)
	Negative	FN (False Negative)	TN (True Negative)

Gambar 3. Confusion Matrix

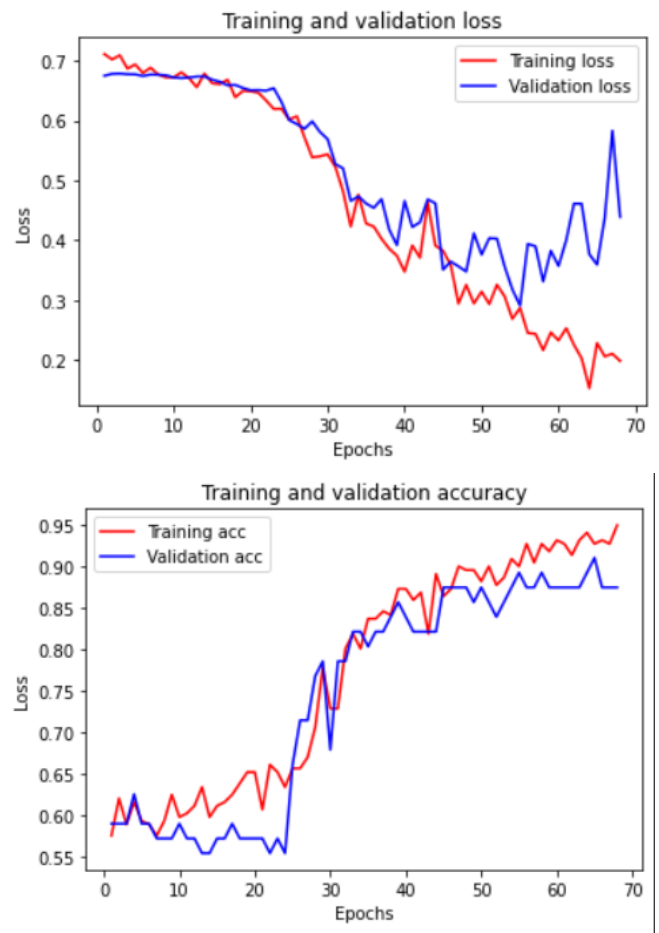
merupakan layer GRU dengan jumlah input sebanyak 40 dan neuron sebanyak 64. Setelah itu diikuti oleh beberapa hidden layer berupa dropout, activation dan Dense.

D. Fitting Data

Sebelum melakukan proses fitting, data dibagi kedalam data training dan juga data testing dengan persentasi 80% dan 20%. Model kemudian dilatih dengan epoch sebanyak 100 kali dan callback untuk mencegah overfitting yaitu jika accuracy sudah diatas 95%. Hasilnya diperoleh 68 kali data melakukan training hingga akhirnya proses ini berhenti.

```
Epoch 68/100: 7/7 [-----] - 0s 26ms/step - loss: 0.1731 - accuracy: 0.9663 - val_loss: 0.4391 - val_accuracy: 0.8750
```

Berikut kurva dari accuracy dan juga loss selama proses fitting:



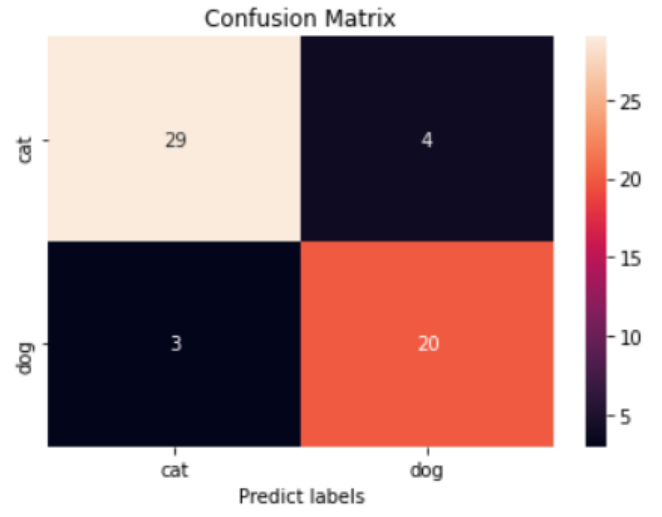
E. Hasil Klasifikasi

Untuk mengukur kinerja model, dilakukan beberapa cara. Diantaranya menggunakan confusion matrix, membuat classification report dan memvisualiasi kurva roc aoc dari model.

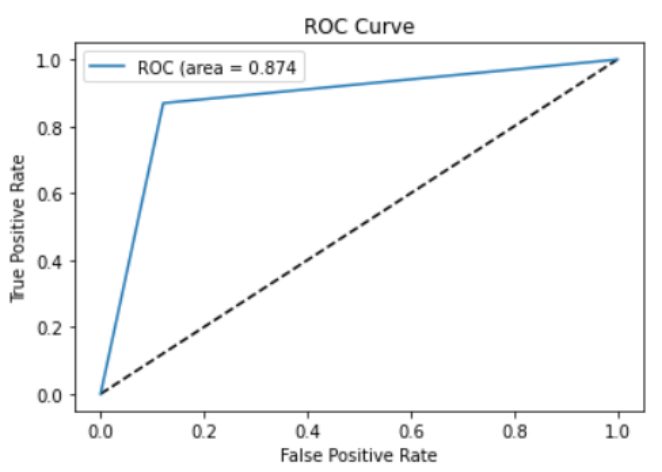
Dengan menggunakan confussion matrix kita dapat melihat berapa banyak data yang diprediksi dengan benar dan berapa banyak data yang diprediksi salah. Terlihat dari confussion matrix ini, 29 data kucing berhasil diprediksi sebagai kucing dan 4 datanya diprediksi salah yaitu sebagai anjing. Berikutnya 20 data anjing berhasil diprediksi dengan benar dan 3 data sisanya diprediksi sebagai kucing.

Dengan menggunakan persamaan accuracy yang dijelaskan sebelumnya diperoleh accuracy sebesar 87.5%. diperoleh pula recall dan presicion berturut-turut sebesar 87% dan 87.5%.

Kemudian divisualisasikan kurva ROC AUC dari hasil



	precision	recall	f1-score	support
0	0.91	0.88	0.89	33
1	0.83	0.87	0.85	23
accuracy			0.88	56
macro avg	0.87	0.87	0.87	56
weighted avg	0.88	0.88	0.88	56



klasifikasi di bawah ini.

V. KESIMPULAN

Dari hasil penelitian ini yaitu pengklasifikasian suara anjing dan kucing menggunakan Long Short Term Memory-Gated Recurrent Unit dan Artificial Neural Network-Back Propagation diambil kesimpulan bahaw LSTM-GRU dianggap baik dalam mengklasifikasikan data

audio anjing dan kucing. Hal ini terbukti dengan nilai akurasi sebesar 87.5%.

UCAPAN TERIMA KASIH

Para penulis mengucapkan terima kasih kepada Tuhan Yang Maha Esa yang telah melimpahkan rahmat-Nya sehingga penulis dapat menyelesaikan penelitian ini. Penulis juga mengucapkan terima kasih kepada Bapak Armin Lawi

dan Khawaritzmi Abdallah Ahmad yang telah membimbing penulis dalam mengerjakan penelitian ini.

DAFTAR ACUAN

- [1] Y. R. Pandeya, D. Kim, and J. Lee, "Domestic cat sound classification using learned features from deep neural nets," *Appl. Sci.*, vol. 8, no. 10, pp. 1–17, 2018.
- [2] L. Nanni, S. Brahnam, A. Lumini, and V. Gradenigo, "Animal Sound Classification Using Dissimilarity Spaces", *Appl.Sci*, vol.10, pp.1-18, 2020.