

KLASIFIKASI GENDER BERDASARKAN SUARA MENGUNAKAN RECURRENT NEURAL NETWORK (RNN)

Diva Tifanny Adherda¹⁾, Missi Hikmatyar²⁾, dan Ruuhwan³⁾

^{1, 2, 3)} Teknik Informatika, Fakultas Teknik, Universitas Perjuangan

^{1, 2, 3)} Jl. Peta No.177, Kahuripan, Kec. Tawang, Kota Tasikmalaya, Jawa Barat 461115 (0265) 326058
e-mail:1903010018@unper.ac.id¹⁾, missi @unper.ac.id²⁾, ruuhwan@unper.ac.id³⁾

Abstrak: Bidang teknologi informasi terus menerus terjadi perkembangan yang pesat. Perkembangan teknologi tidak lepas dari beberapa faktor seperti sentuhan, penglihatan maupun suara. Setiap manusia dengan manusia yang lain memiliki perbedaan karakteristik, salah satunya yang dapat dilihat yaitu dari suaranya. Pemrosesan suara adalah konsep yang sangat penting untuk semua jenis sistem yang membutuhkan interaksi manusia dalam aktivitas sehari-hari. Adapun salah satu teknik yang digunakan dalam pemrosesan ucapan yaitu klasifikasi, yang berdampak langsung pada sistem pengenalan ucapan. SimpleRNN dan LSTM adalah model deep learning yang bisa dipakai untuk mengklasifikasikan sentimen. Metode ini bisa mengolah data dengan berurutan seperti suara, video, dan teks. Hasil penelitian ini mampu memberikan akurasi 90% pada data uji dan akurasi 95% pada data latih.

Kata Kunci— gender, klasifikasi, pengenalan suara, RNN

Abstract: The information technology field continues to progress rapidly. Technological progress is kept in check with such factors as touch, sight, and sound. Each man with another man has a characteristic difference, one that can be seen is by his voice. The processing of sound is an essential concept to all kinds of systems that require human interaction in its daily activities. One of the techniques used in processing speech is classification, which has a direct effect on speech recognition systems. SimpleRNN and LSTM are models of deep learning that can be used to classify sentiment. It can process data in such a sequence as sound, video, and text. These results provide accuracy 90% of the test data and 95% accuracy to the training data.

Keywords— gender, classification, voice recognition, RNN

I. PENDAHULUAN

Bidang teknologi informasi terus menerus terjadi perkembangan yang pesat. Perkembangan teknologi tidak lepas dari beberapa faktor seperti sentuhan, penglihatan maupun suara. Setiap manusia dengan manusia yang lain memiliki perbedaan karakteristik, salah satunya yang dapat dilihat yaitu dari suaranya atau istilah lainnya yaitu biometrik. Biometrik merupakan ilmu mengenai pengenalan suatu ciri khas manusia dari keunikannya. Setiap manusia mempunyai keunikan yang berbeda dengan yang lainnya. Selain ciri khas dari bentuk wajah, DNA, sidik jari, dan iris, ada keunikan lain yang termasuk ciri khas dari manusia yaitu suara [1].

Dalam kehidupan sehari-hari, manusia dapat mengenali suara dengan mudah, yaitu dengan cara mendengarkan dengan penuh pemahaman yang diucapkan oleh manusia, dan manusia juga mempunyai kecerdasan dalam mengenali pola suara. Namun, jika ada ribuan bahkan jutaan suara untuk diklasifikasi, akan menyulitkan manusia, sehingga diperlukan sistem yang dapat mengklasifikasikan suara secara otomatis. Tidak seperti kasus dalam komputer, proses pengenalan suara merupakan proses yang sulit, karena komputer memerlukan mekanisme logika standar untuk mengenali pola suara [2].

Suara laki-laki dan perempuan mempunyai masing-masing karakteristik, karena getaran yang berbeda di tenggorokan, dan dengan memproses sinyal suara, karakteristik ini akan diperoleh dari sesuatu yang dapat dikenal oleh sistem. Dengan fitur ini, sistem bisa melakukan identifikasi *gender* dari sinyal suara [3]. Suara yang diolah termasuk hal yang penting bagi sebuah sistem yang memerlukan interaksi dari manusia. Teknik yang dipakai dalam pengolahan suara yaitu klasifikasi dengan pengaruh langsung suatu sistem pengenalan suara [4]. Kendala-kendala yang mungkin terjadi saat proses

ekstraksi adalah suara saat kondisi seorang sakit, dan lingkungan, contohnya seperti efek mikrofon, kebisingan, dan gema. Maka dari itu, pada tahap preprocessing diperlukan proses penghilangan noise.

Pada penelitian yang dilakukan [5] dapat ditarik kesimpulan, pembuatan sistem untuk mengenali ucapan Bahasa Indonesia menggunakan MFCC dan RNN, menghasilkan rata-rata akurasi tertinggi yaitu 73.55%. Adapun penelitian lain yang dilakukan [6] dengan judul Klasifikasi suara berdasarkan *gender* dengan metode *K-Nearest Neighbor (KNN)* memberikan akurasi rata-rata terendah sebesar 71,5% dan tertinggi sebesar 76,2%. Selanjutnya, pada penelitian yang dilakukan [7] dengan judul klasifikasi penyakit jantung menggunakan metode *RNN* memberikan nilai akurasi model klasifikasi terbaik 4% pada setiap kasus. Yaitu untuk kasus data yang tidak seimbang, menghasilkan akurasi pelatihan sebesar 99,81% dan akurasi uji sebesar 99,81%. Sedangkan untuk data seimbang, akurasi pelatihan adalah 99,98% dan akurasi tes adalah 99,96%. Selain itu pada penelitian yang dilakukan [8] dengan judul klasifikasi *gender* berdasarkan suara dengan metode *naïve bayes* dan MFCC menghasilkan tingkat akurasi keberhasilan yaitu sebesar 87% dari jumlah data 100 sampel.

Metode *Recurrent neural network (RNN)* adalah metode yang dapat dengan baik mengenal ucapan, karena kemampuannya yaitu mengenali suatu informasi dengan berbagai varian seperti halnya sinyal ucapan. RNN memiliki kemampuan dalam menyimpan memori atau *feedback loop* yang memungkinkan mengenali suatu pola data [2]. Pada tugas pengklasifikasian suara, SimpleRNN dapat digunakan untuk mengambil urutan fitur suara sebagai input dan mengklasifikasikannya ke dalam kategori tertentu. Adapun LSTM dirancang khusus untuk mengatasi masalah-masalah dengan memori jangka panjang dalam deret waktu. Pintu-pintu (*gate*) dalam LSTM memungkinkan jaringan untuk mengatur aliran informasi dan mengontrol memori jangka panjang.

Tujuan penelitian adalah menentukan tingkat akurasi dengan melakukan klasifikasi *gender* dengan menggunakan metode *RNN* yang diharapkan dapat bermanfaat untuk keperluan terkait dengan hal-hal yang membutuhkan klasifikasi suara. Dengan metode ini, diharapkan sistem mampu mengenali suara berdasarkan gender yang menghasilkan akurasi dengan tingkatan yang baik dan optimal.

II. TINJAUAN PUSTAKA

A. Deep Learning

Deep learning merupakan hal yang sangat penting dalam bidang *AI* yang menggunakan banyak lapisan dalam pemrosesan data non-linier yang diimplementasikan dalam *fitur extraction*, klasifikasi dan pengenalan pola. *Deep learning* yaitu metode untuk menangani masalah dalam kerangka pembelajaran komputer yang menggunakan konsep *chain of interest*. Suatu ide rantai dapat memperdayakan komputer untuk mempelajari berbagai ide kompleks dengan menggabungkan berbagai ide yang kurang kompleks. Maka dari itu, hal ini bisa membantu untuk menghasilkan pemahaman yang lebih mendalam terkait proses *deep learning* itu sendiri berdasarkan banyak lapisan yang telah digambarkan dalam grafik [9].

B. Klasifikasi

Klasifikasi adalah suatu proses penganalisisan data dengan membentuk suatu model untuk mendeklarasikan objek ke dalam kategori yang telah ditentukan. Dalam klasifikasi data terdiri dari dua tahapan, yaitu yang pertama adalah pelatihan atau pembelajaran (*learning*), yaitu algoritma klasifikasi melatih sebuah model dengan menganalisa kumpulan data pelatihan. Pada tahap selanjutnya, model yang dilatih, dipakai sebagai pengklasifikasian data uji. Akurasi pengklasifikasian data uji adalah presentase data uji yang diklasifikasikan [10].

C. Pengenalan Suara

Suara manusia adalah getaran yang dihasilkan dari pita suara manusia dan dapat didengar oleh manusia itu sendiri. Getaran yang dimaksudkan ini dapat berupa suara berbicara, suara berteriak, suara menangis, dan suara yang lainnya. Pengenalan suara adalah proses mengidentifikasi dan memahami pola-pola atau ciri-ciri dalam sinyal suara untuk mengklasifikasikan atau menerjemahkan suara ke dalam bentuk yang dapat dimengerti oleh komputer atau sistem. Tujuan utamanya adalah untuk mengubah sinyal suara menjadi informasi yang berguna, seperti teks, perintah, atau tindakan tertentu. Berdasarkan pembagian klasifikasi frekuensi suara, frekuensi suara dibagi menjadi 3 jenis, yaitu

infrasonik yang mana frekuensi suara dibawah 20 Hz, kemudian frekuensi audiosonik yang mana frekuensi tersebut getaran suaranya berada diantara 20 Hz sampai dengan 20000 Hz dan yang terakhir adalah frekuensi ultrasonik, dimana frekuensi ini getaran suaranya berada lebih dari 20000 Hz [8].

D. Gender

Klasifikasi *gender* dilakukan terhadap dua jenis kelas yaitu perempuan dan laki-laki. Suara yang dihasilkan dari kedua *gender* memiliki perbedaan berdasarkan bentuk tenggorokannya. Perbedaan suara dapat diketahui dengan menggunakan teknik pemrosesan sinyal audio yang tepat. Banyak metode telah disarankan untuk membangun suatu sistem klasifikasi *gender* berbasis suara berkualitas tinggi, tetapi secara umum, masih kurang efisien terhadap derau atau noise [11].

1. Laki-laki



Gambar 1. Gelombang Suara Laki-Laki

2. Perempuan



Gambar 2. Gelombang Suara Perempuan

E. Recurrent neural network (RNN)

Metode RNN dapat bekerja dengan baik dalam mengenali ucapan karena dengan kemampuannya mengenali informasi dengan banyaknya varian waktu seperti sinyal ucapan. RNN merupakan jaringan dengan minimum satu *feedback loop*, yaitu *recurrent network* mempunyai satu *layer* neuron tunggal yang masing-masing memberikan *feedback* output-nya sebagai masukan dari keseluruhan *neuron*. RNN memiliki kemampuan menyimpan memori atau *feedback loop* yang dapat mengenali pola data dengan baik [5].

Pada proses RNN (*Recurrent Neural Network*) data akan berulang kali dipanggil untuk memproses input, yaitu disebut juga dengan data *sequential*. Proses klasifikasi deret waktu dan data sekuensial adalah fitur ciri khasnya. Data time series adalah data yang disatukan secara kronologis dari interval tertentu, sedangkan data sekuensial adalah data sampel yang dilakukan proses dengan berurutan yang di mana setiap deret saling berhubungan.

F. SimpleRNN

SimpleRNN merupakan suatu tipe arsitektur jaringan saraf buatan yang digunakan dalam pembelajaran mesin untuk tugas yang melibatkan data urutan, seperti *teks* atau seri waktu. Arsitektur ini merupakan salah satu bentuk dasar dari *RNN* dan biasanya digunakan untuk masalah yang melibatkan urutan data yang pendek atau sederhana. *SimpleRNN* memiliki kemampuan untuk "mengingat" informasi sebelumnya dalam urutan data dan menggunakan informasi tersebut untuk membuat prediksi atau analisis terhadap data selanjutnya dalam urutan tersebut. Namun, satu masalah dengan *SimpleRNN*

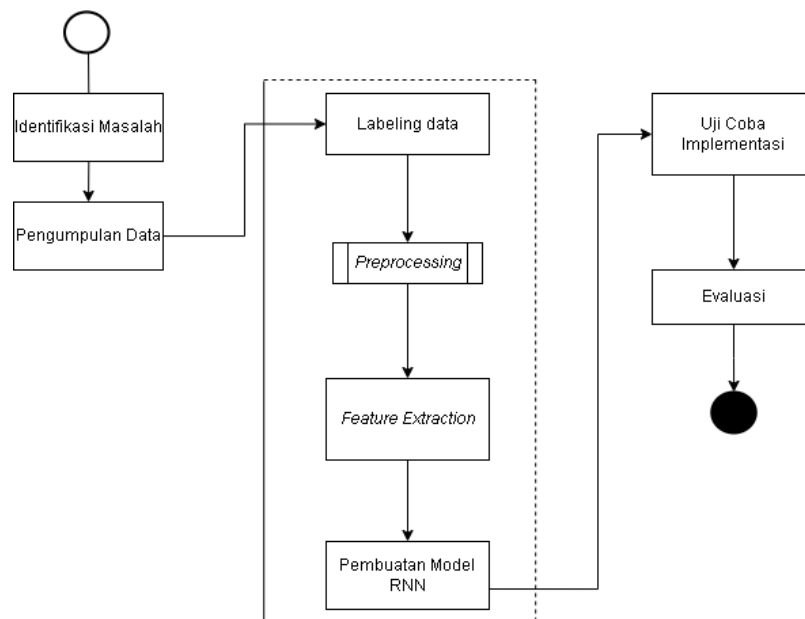
adalah kemampuannya untuk mengatasi masalah *vanishing gradient*, yang dapat membuatnya sulit untuk menangkap ketergantungan jangka panjang dalam data urutan yang panjang.

G. Long Short Term Memory (LSTM)

LSTM adalah salah satu modifikasi dari recurrent neural network atau *RNN*. *LSTM* adalah jenis arsitektur jaringan saraf rekuren (*RNN*) yang dirancang khusus untuk mengatasi masalah masalah *vanishing gradient* dan mengatasi ketergantungan jarak jauh dalam data urutan panjang. *LSTM* adalah pengembangan lebih lanjut dari jaringan saraf rekuren konvensional, seperti *SimpleRNN*, dengan kemampuan untuk "mengingat" informasi jangka panjang dari data urutan yang masuk.

III. METODE PENELITIAN

Untuk memberikan gambaran mengenai proses pada pengerjaan penelitian ini. Berikut adalah alur metodologi penelitian yang peneliti buat.



Gambar 3. Metode Penelitian

A. Identifikasi Masalah

Pada tahap awal penelitian, peneliti melakukan identifikasi permasalahan berdasarkan penelitian-penelitian yang berkaitan dengan deep learning, artificial intelligent dan Metode *Recurrent Neural Network*. Pada penelitian yang sudah ada maupun yang belum pernah ada yang membahas permasalahan pada sistem *voice recognition* dari sisi belakang dan juga atas dengan menggunakan *deep learning*.

B. Pengumpulan Data

Langkah pengumpulan data dilakukan dengan berbagai cara untuk mengumpulkan informasi yang dibutuhkan, yaitu:

1. Studi literatur yaitu suatu metode pengumpulan data dengan melakukan *review* dan membandingkan literatur yang berasal dari jurnal, buku, dan juga hasil penelitian sebelumnya.
2. Data sampel berjumlah 100 file suara berformat mp3. Sampel suara memuat 50 suara laki-laki dan 50 suara perempuan. Data tersebut diperoleh dari rekaman suara pada *handphone* dengan durasi berkisar antara 3-7 detik.

C. Labeling Data

Labeling data bertujuan untuk membuat data pengujian dan data pelatihan. Pelabelan dilakukan pada data yang belum ada label jenis kelaminnya. Pelabelan dilakukan secara subjektif. Nama-nama tersebut kemudian dinotasikan dengan dua label yaitu L untuk *gender* laki-laki dan P untuk *gender* perempuan.

D. Preprocessing

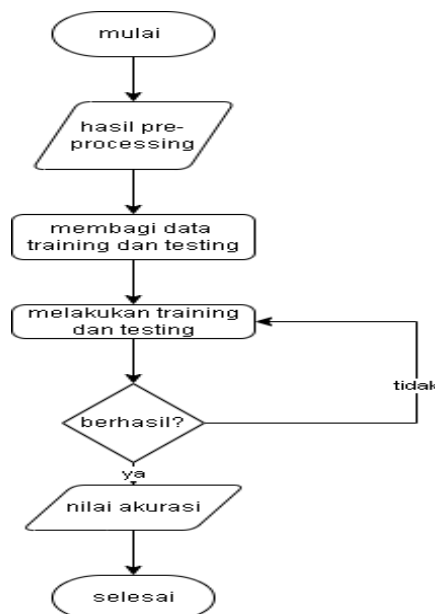
Proses ini yaitu tahap yang mentransformasi data dari yang bentuknya tidak terstruktur menjadi data yang terstruktur. Preprocessing adalah proses untuk menghilangkan bagian yang tidak dibutuhkan untuk perubahan data yang berkualitas yaitu data yang telah memenuhi syarat untuk dilakukan eksekusi. Bentuk pembersihan data ini yaitu dengan cara menghapus data suara yang *double* dan menghilangkan *noise* pada sampel suara. Data yang sudah terstruktur tersebut kemudian akan dianalisis untuk menentukan polaritas suara apakah laki-laki, atau perempuan. Setelah data di preprocessing, data menjadi dataset yang siap untuk diproses. Lalu data dipecah menjadi 2, data latih dan data uji.

E. Feature Extraction

Proses perubahan data menjadi representasi yang lebih sederhana dan informatif yang dapat digunakan untuk mewakili karakteristik penting dari data tersebut. Ketika data memiliki banyak fitur (variabel) atau fitur-fiturnya memiliki dimensi yang tinggi, ekstraksi fitur membantu untuk meringkas informasi ke dalam representasi yang lebih ringkas dan lebih mudah diolah.

F. Pembuatan Model RNN

Tahapan ini yaitu melakukan proses rancangan klasifikasi dengan metode RNN:



Gambar 4. Flowchart Model Recurrent Neural Network (RNN)

G. Uji Coba Implementasi

Uji coba implementasi merupakan tahap pengklasifikasian dilakukan satu per satu sesuai banyaknya sampel dan dilakukan berulang sesuai dengan metode RNN. Uji coba ini dilakukan untuk memperlihatkan bahwa pengklasifikasian telah berjalan dengan semestinya yaitu input dataset, proses dan output sesuai dengan kebutuhan.

H. Evaluasi

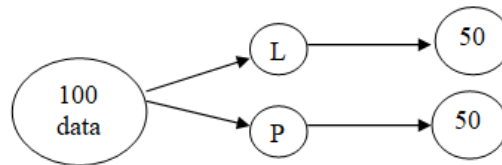
Pada tahap ini merupakan proses untuk mengukur dan menganalisis kinerja sebuah model atau sistem yang bertujuan untuk mengklasifikasikan suara ke dalam beberapa kategori atau kelas yang telah ditentukan sebelumnya. Proses evaluasi ini bertujuan untuk menilai sejauh mana model atau sistem yang dibangun dapat mengenali dan membedakan berbagai kelas suara dengan akurat.

IV. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Pada tahap ini merupakan proses memperoleh data audio dari sumber audio secara real time Data diperoleh dari rekaman suara pada *Handphone* dengan durasi berkisar antara 3-7 detik dengan format

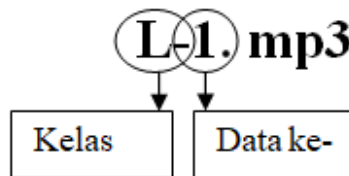
.mp3. Adapun faktor-faktor yang dipertimbangkan dalam pengumpulan data yaitu kualitas suara, jumlah sampel, dan keseimbangan data. Terdiri dari 100 pembicara yaitu 50 laki-laki dan 50 perempuan yang diminta untuk mengucapkan kalimat yang telah ditentukan. Rekaman suara pria diberi kode "L", sedangkan rekaman suara wanita diberi kode "p".



Gambar 5. Pengumpulan Data

B. Labeling Data

Dalam konteks pengenalan gender menggunakan suara, label data mengidentifikasi apakah suara yang dikaitkan dengan setiap sampel adalah suara laki-laki atau perempuan dengan cara memberikan tag atau kategori pada data suara yang sudah dikumpulkan. Setiap sampel suara akan diberi label yang sesuai dengan jenis kelamin yang terkait dengan suara tersebut.



Gambar 6. Contoh Penamaan Pada Tiap-tiap Dataset

Adapun jika di representasikan pada semua data dapat ditunjukkan pada tabel berikut

Tabel 1. Pembagian label pada tiap kelas

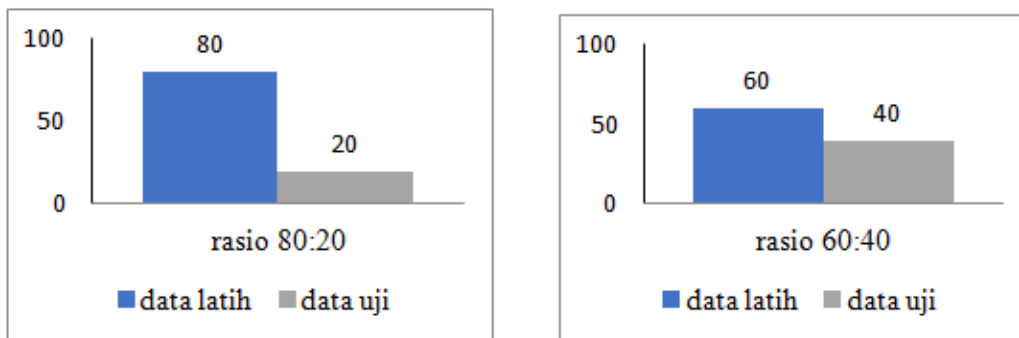
Kelas	Label	Jumlah Data
Laki-laki	0	50
Perempuan	1	50
Jumlah Data		100

C. Preprocessing

Tahap pra-pemrosesan (*pre-processing*) dalam pengenalan gender menggunakan suara yaitu tahap di mana data suara dari dataset dipersiapkan sebelum masuk ke tahap ekstraksi fitur atau pelatihan model. Tahap *Preprocessing* ini bertujuan untuk mengurangi kebisingan atau suara yang tidak relevan pada data suara yang telah dikumpulkan menggunakan website penghilang noise online pada link <https://noisereducer.media.io/speech-enhancement>.

Pada tahap *preprocessing*, Dataset yang digunakan dibagi menjadi dua bagian pada proses data splitting, yaitu:

- Data latih: dipakai untuk melatih model pengenalan pola dalam data audio dan memetakan pola tersebut ke label gender yang sesuai. Data pelatihan terdiri dari beberapa rekaman suara dari jenis kelamin yang berbeda.
- Data uji: dipakai untuk mengevaluasi kinerja model yang telah dilatih.



Gambar 7. Rasio Data latih dan data uji 80:20 dan 60:40

Pada gambar 7, memuat gambaran diagram yang merupakan perbandingan dua rasio yaitu yang pertama 80 data latih, dan 20 data uji, kemudian yang kedua yaitu 60 data latih dan 40 data uji. Perbandingan rasio tersebut bertujuan untuk memastikan dan mendapatkan perkiraan bahwa evaluasi model lebih objektif tentang seberapa baik performa dari model yang di uji atau dilatih.

D. Feature Extraction

Pada proses *feature extraction* mengacu pada proses mengubah sinyal suara mentah menjadi representasi yang lebih kompak dan informatif yang dapat digunakan oleh algoritma pengenalan suara atau model pembelajaran mesin untuk memahami ciri-ciri penting dari suara tersebut. Tujuan proses *feature extraction* adalah untuk mengurangi dimensi data suara dan menyoroti informasi yang paling relevan dan membedakan dari sinyal suara. Sinyal suara disini berbentuk data kontinu dalam bentuk gelombang.

```
def extract_feature(file_name):  
    X, sample_rate = librosa.load(file_name)  
    stft=np.abs(librosa.stft(X))  
    result=np.array([])  
    mfccs=np.mean(librosa.feature.mfcc(y=X, sr=sample_rate, n_mfcc=40).T,axis=0)  
    result=np.hstack((result, mfccs))  
    chroma=np.mean(librosa.feature.chroma_stft(S=stft, sr=sample_rate).T,axis=0)  
    result=np.hstack((result, chroma))  
    mel=np.mean(librosa.feature.melspectrogram(y=X, sr=sample_rate).T,axis=0)  
    result=np.hstack((result, mel))  
    return result
```

Gambar 8. Feature Extraction

E. Pembuatan Model RNN

Pemodelan dilakukan dengan menggunakan *framework Tensorflow* dan *library SimpleRNN*. *Library* menyesuaikan dengan proses pembangunan model dan mampu menentukan parameter dari berbagai komponen pendukung model sehingga model memperoleh performa yang optimal dalam memahami data uji dan latih yang diberikan sebagai input. Dalam pengujian ini, *SimpleRNN* dan *LSTM* dipanggil dengan parameter satuan 64. Hal tersebut dibuat dengan mempertimbangkan secara komprehensif (menyeluruh) akurasi pelatihan, akurasi tes, waktu pelatihan, dan menghindari *overfitting*. Untuk pemodelan pada *code* nya dapat dilihat pada gambar 9

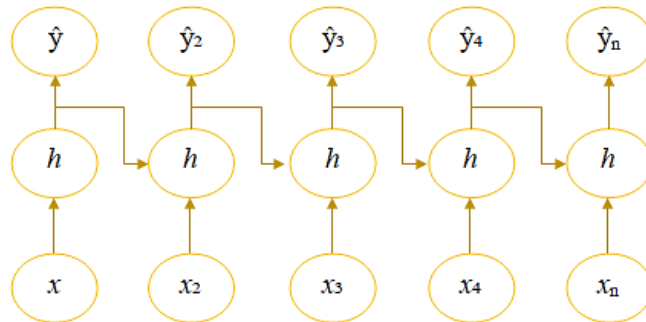
```
model = Sequential()  
model.add(SimpleRNN(64, input_shape=(180, 1), return_sequences=True)),  
tf.keras.layers.Dropout(0.2)  
model.add(LSTM(64))  
model.add(Dense(2, activation='sigmoid'))
```

Gambar 9. Source Code Pemodelan RNN

Pada dasarnya untuk klasifikasi biner dengan dua kelas bisa menggunakan aktivasi *softmax* maupun *sigmoid*. Keduanya dapat digunakan sebagai fungsi aktivasi pada *layer output* dalam model. Alasan

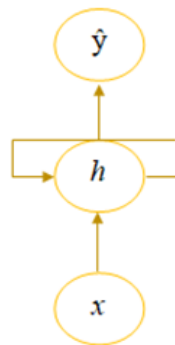
penulis lebih menggunakan aktivasi *sigmoid* karena fungsi aktivasi tersebut biasanya digunakan dalam klasifikasi biner karena menghasilkan output antara 0 dan 1. Output yang dihasilkan dapat diinterpretasikan sebagai probabilitas kelas positif. Selain itu, jika penulis menggunakan aktivasi *softmax* maka nantinya akurasi akan cenderung lebih kecil dibandingkan dengan *sigmoid*.

Pada tahap pembuatan model ini nantinya RNN berulang kali melakukan perhitungan yang sama untuk input yang diberikan.



Gambar 10. Ilustrasi Proses pengulangan RNN

Arsitektur tersebut bisa di sederhanakan dengan bentuk pada gambar seperti Gambar 11.

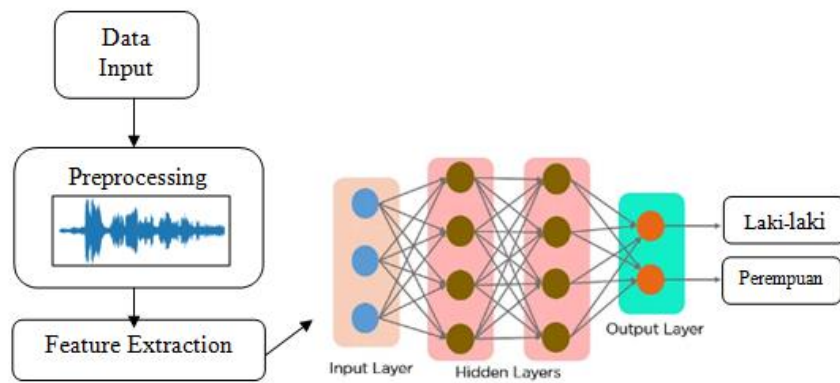


Gambar 11. Arsitektur Sederhana RNN

Pada gambar diatas menjelaskan proses dari RNN. Jaringan ini sedikit berbeda dari RNN biasa. Jika di RNN semua input masuk dan melewati lapisan tersembunyi lalu mengeluarkan *output*, tetapi di RNN ada proses *loop*. Setiap masukan yang masuk dan menghasilkan keluaran dimasukkan kembali sebagai masukan untuk diproses di lapisan tersembunyi dan seterusnya agar output sesuai dengan target [16].

Menurut [11] Kelemahan model menggunakan arsitektur *simpleRNN* dapat diatasi dengan cara menutupi keterbatasan model melalui penggabungan model dengan arsitektur berbeda. Model dengan kelebihan pada sisi yang menjadi kelemahan *simpleRNN* diperhitungkan sebagai pasangan arsitektur yang cocok dalam pembuatan *Hybrid Model*. Pada penelitian [12] memanfaatkan penggabungan arsitektur *simpleRNN* dan LSTM untuk memprediksi masa sisa pakai turbin angin, LSTM memiliki kemampuan yang baik dalam memprediksi data dengan ketergantungan jangka panjang (*Long Term*). Hasil dari penelitian tersebut menunjukkan bahwa *Hybrid Model* mampu mengatasi kelemahan arsitektur *simpleRNN*, model mampu mengatasi ketergantungan jangka panjang dan memberikan performa yang baik pada hasil evaluasi.

Adapun alur dalam memproses data suara menjadi *output* dapat ditunjukkan pada gambar berikut:



Gambar 122. Alur proses pemodelan RNN

Penting untuk memilih jumlah *epoch* yang tepat saat melatih model. Maka dari itu, pada penelitian ini digunakan perbandingan *epoch*. *Epoch* adalah jumlah kali keseluruhan dataset diberikan ke dalam model selama proses pelatihan. Setiap *epoch* melibatkan proses penggunaan data latih untuk melakukan pembelajaran dan penyesuaian bobot dan parameter model. Pada awal pelatihan, model mungkin belum mempelajari pola dan hubungan dalam data pelatihan dengan benar. Selama masa ini, model secara bertahap meningkatkan kinerja dan meningkatkan akurasi. Namun, salah satu faktor yang diperhatikan yakni kemungkinan terjadinya *underfitting* dan *overfitting*. *Epoch* yang terlalu sedikit dapat menyebabkan *underfitting*, di mana model gagal mempelajari pola kompleks dalam data pelatihan. Di sisi lain, terlalu banyak *epoch* dapat menyebabkan *overfitting*, di mana model "menghafal" data pelatihan dengan baik, tetapi gagal menggeneralisasi data baru.

Pada penelitian kali ini, dilakukan pelatihan model terhadap data latih Tabel 2.

Tabel 2. Hasil Akurasi Data Uji

Data Uji (%)	Epoch				
	50	100	150	200	250
20	0.60	0.80	0.60	0.90	0.75
40	0.77	0.67	0.57	0.65	0.60

Dapat dilihat bahwa hasil pengujian akurasi menggunakan perbandingan data uji 20% dan 40% serta perbandingan lima *epoch* antara 50-250. Hasil terbaik yang didapatkan pada model RNN adalah pada persentase data uji 20% dan data latih 80% yang terletak pada epoch 200.

Accuracy Training: 0.949999988079071
 Accuracy Testing: 0.8999999761581421

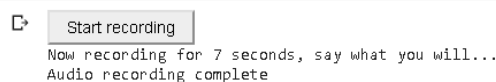
Gambar 13. Hasil Akurasi Data Latih dan Data Uji

Pada hasil tersebut didapatkan nilai akurasi tertinggi yaitu data latih sebesar 95% dan data uji sebesar 90%.

F. Uji Coba Implementasi

Setelah model berhasil dibuat, proses berlanjut ke tahap implementasi. Proses ini dilakukan dengan membuat dua fungsi yaitu:

- Fungsi rekam. Tujuan dari fitur ini adalah untuk menangkap *audio* langsung dari pengguna dan mengubahnya menjadi file *audio*.



Gambar 14. Proses Perekaman Suara

b. Fungsi identifikasi. Fungsi ini bertujuan untuk mengolah *file audio* yang direkam melalui proses ekstraksi data, sehingga dapat dijadikan input untuk digunakan oleh model. Fungsi ini menampilkan hasil pengenalan apakah suara tersebut termasuk dalam kategori Laki-laki atau Perempuan.

```
1/1 [=====] - 0s 41ms/step
The system detects that you are a Laki-Laki
```

Gambar 15. Proses Identifikasi Suara

G. Evaluasi

Untuk membantu dalam evaluasi kinerja model klasifikasi dengan memberikan gambaran tentang sejauh mana model dapat memprediksi dengan benar dan sejauh mana model mengalami kesalahan dalam melakukan klasifikasi. Maka perlu untuk mengetahui hasil performa model *Recurrent Neural Network* (RNN), dengan adanya tahapan evaluasi. Parameter yang digunakan untuk performansi klasifikasi yaitu akurasi, *precision*, *recall* dan *F1-score*.

	precision	recall	f1-score	support
0	0.80	1.00	0.89	8
1	1.00	0.83	0.91	12
accuracy			0.90	20
macro avg	0.90	0.92	0.90	20
weighted avg	0.92	0.90	0.90	20

Gambar 16. Confusion Matrix

1. Accuracy

Accuracy merupakan proses pengukuran sejauh mana model dapat membuat prediksi yang benar secara keseluruhan. Ini dinyatakan sebagai rasio prediksi yang benar (*True Positive* dan *True Negative*) terhadap keseluruhan jumlah data. Untuk mencari nilai *accuracy* yaitu:

$$\begin{aligned}
 \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} & (1) \\
 &= \frac{8+10}{8+10+0+2} \\
 &= \frac{18}{20} \\
 &= 0,9 * 100\% \\
 &= 90\%
 \end{aligned}$$

2. Precision

Precision merupakan proses mengukur sejauh mana prediksi positif model adalah benar. Dinyatakan sebagai rasio *True Positive* terhadap total prediksi positif.

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{TP+FP} & (2) \\
 &= \frac{8}{8+0} \\
 &= 1.00 \\
 &= 100\%
 \end{aligned}$$

3. Recall

Recall yaitu proses mengukur sejauh mana model mampu mengidentifikasi semua *instance* positif yang sebenarnya. Dinyatakan sebagai rasio *True Positive* terhadap total *instance* positif (*True Positive* dan *False Negative*). Untuk mencari nilai *recall* yaitu:

$$\text{Recall} = \frac{TP}{TP+FN} \tag{3}$$

$$\begin{aligned} &= \frac{8}{8+2} \\ &= 0,8 \\ &= 80\% \end{aligned}$$

4. *F1-Score*

F1-score merupakan harmonisasi antara *precision* dan *recall*. Ini memberikan keseimbangan antara ketepatan dan kelengkapan model. *F1-Score* lebih berguna ketika ada ketidakseimbangan antara kelas positif dan negatif. *F1-score* juga menjadi salah satu metrik evaluasi klasifikasi yang digunakan untuk mengukur performa model dalam memprediksi kelas yang benar. Nilai terbaik *F1-Score* adalah 1.0 dan nilai terburuknya adalah 0. Secara representasi, jika *F1-Score* punya skor yang baik mengindikasikan bahwa model klasifikasi punya *precision* dan *recall* yang baik. Untuk mencari nilai *F1-score* yaitu:

$$\begin{aligned} F1 - Score &= 2 \times \frac{Recall \times precision}{Recall + precision} & (4) \\ &= 2 \times \frac{0,8 \times 1,0}{0,8+1,0} \approx 0,88 \end{aligned}$$

V. KESIMPULAN DAN SARAN

KESIMPULAN

Pada hasil dan pembahasan penelitian ini, dapat ditarik kesimpulan:

1. Melakukan klasifikasi *gender* berdasarkan suara dari 100 *dataset* yaitu 50 data suara laki-laki, dan 50 data suara perempuan menggunakan metode *Recurrent Neural Network* (RNN) dengan arsitektur *simpleRNN* dan *LSTM*.
2. Mengimplementasikan dan mengklasifikasikan *gender* berdasarkan suara dengan metode *Recurrent neural network* (RNN) yang dibangun mampu memberikan akurasi data uji sebesar 90% dan data latih 95% dengan perbandingan rasio 20:80 dan *epoch* 200.

SARAN

Peneliti mengajukan saran untuk penelitian ini lebih baik kedepannya, termasuk menambahkan *dataset* dan mengkombinasikan metode untuk mendapatkan hasil akurasi yang lebih optimal.

DAFTAR PUSTAKA

- [1] N. Izzah, “KLAUSTERING SUARA BERDASARKAN GENDER MENGGUNAKAN ALGORITMA K-MEANS DARI HASIL EKSTRAKSI FFT (Fast Fourier Transform),” *J. Ilm. Soulmath J. Edukasi Pendidik. Mat.*, vol. 6, no. 1, pp. 47–58, 2018, doi: 10.25139/sm.v6i1.790.
- [2] A. Apsarini, “Klasifikasi Jenis Kelamin Berdasarkan Suara Menggunakan Metode Learning Vector,” vol. 4, no. 7, pp. 2301–2308, 2020.
- [3] D. A. Adi Rinaldi, Hendra, “Pengenalan Gender Melalui Suara dengan Algoritma Support Vector Machine (SVM),” no. x, pp. 1–10, 2016.
- [4] Y. Mustaqim, E. Utami, and S. Raharjo, “KLASIFIKASI AUDIO MENGGUNAKAN WAVELET TRANSFORM DAN NEURAL NETWORK Yulianto,” *J. Inform. Dan Teknol. Inf.*, vol. 4, no. 2, pp. 122–130, 2019.
- [5] P. Tridarma and S. N. Endah, “Pengenalan Ucapan Bahasa Indonesia Menggunakan MFCC dan Recurrent Neural Network,” *J. Masy. Inform.*, vol. 11, no. 2, pp. 36–44, 2020, [Online].

Available: <https://ejournal.undip.ac.id/index.php/jmasif/article/view/34874>

- [6] D. E. R. Riska Yessivirna 1 , Marji2, “Klasifikasi Suara Berdasarkan Gender (Jenis Kelamin) Dengan Metode K-Nearest Neighbor (Knn),” *Univ. Brawijaya Malang*, no. 37, pp. 1–31, 2015.
- [7] H. Azis, *Klasifikasi Penyakit Jantung Menggunakan Metode Recurrent Neural Network (RNN)*. 2018.
- [8] S. Safriadi and R. Rahmadani, “Klasifikasi Gender Berdasarkan Suara Dengan Naive Bayes Dan Mel Frequency Cepstral Coefficient,” *VOCATECH Vocat. Educ. Technol. J.*, vol. 2, no. 1, pp. 19–26, 2020, doi: 10.38038/vocatech.v2i1.45.
- [9] F. A. Hermawati and R. A. Zai, “Sistem Deteksi Pemakaian Masker Menggunakan Metode Viola-Jones dan Convolutional Neural Networks (CNN),” *Proceeding KONIK (Konferensi Nas. Ilmu Komputer)*, vol. 5, pp. 182–187, 2021.
- [10] R. A. Asmara, B. S. Andjani, U. D. Rosiani, and P. Choirina, “KLASIFIKASI JENIS KELAMIN PADA CITRA WAJAH MENGGUNAKAN METODE NAIVE BAYES,” pp. 212–217, 2018.
- [11] R. B. Handoko and S. Suyanto, “Klasifikasi Gender Berdasarkan Suara Menggunakan Support Vector Machine,” *Indones. J. Comput.*, vol. 4, no. 1, p. 9, 2019, doi: 10.21108/indojc.2019.4.1.244.