

PERANCANGAN SISTEM *VOICE RECOGNITION* BERBASIS AI PADA LAYANAN *DRIVE THRU* RESTORAN CEPAT SAJI

1st Revin Prasya Danendra
Teknik Telekomunikasi
Telkom University Surabaya
Surabaya, Indonesia

revinprasyadanendra@student.telkomuniversity.ac.id

2nd Nilla Rachmaningrum
Teknik Telekomunikasi
Telkom University Surabaya
Surabaya, Indonesia

nrachmaningrum@telkomuniversity.ac.id

3rd Muhammad Adib Kamali
Teknologi Informasi
Telkom University Surabaya
Surabaya, Indonesia

adibmkamali@telkomuniversity.ac.id

Abstrak — Perkembangan bisnis selalu meningkat seiring kebutuhan manusia yang bersifat konsumtif dalam bidang jasa layanan yaitu layanan *drive-thru*. Sistem layanan *drive-thru* pada restoran cepat saji menghadapi tantangan dalam efisiensi dan akurasi pemesanan, terutama untuk mengurangi biaya dan meningkatkan produktivitas layanan *drive-thru*. Penelitian ini bertujuan merancang dan mengaplikasikan sistem *voice recognition* berbasis AI pada layanan *drive thru* restoran cepat saji untuk meningkatkan keakuratan pemesanan. Teknologi *voice recognition* digunakan untuk mengubah gelombang suara menjadi bentuk teks dari model yang telah dilatih. Metode yang digunakan adalah metode Short Time Fourier Transform untuk membangun model Convolutional Neural Network dengan library Tensorflow menggunakan Python. Hasil dari penelitian yaitu akurasi sebesar 98%, presisi sebesar nilai 98%, recall sebesar 98%, dan F1-score sebesar 98%. Serta selama pelatihan 10 epoch mengalami kenaikan pada nilai akurasi dan penurunan pada nilai *loss*. Hal ini menunjukkan bahwa model dapat mengenali pola – pola dan mampu mengeneralisasi data yang baru. Model yang dilatih disimpan dalam format *h5* dan diimplementasikan secara nyata.

Kata kunci— *Voice recognition*, Convolutional Neural Network, *Drive thru*, Tensorflow, Python

I. PENDAHULUAN

Perkembangan industri makanan dan minuman semakin meningkat seiring kebutuhan manusia terutama restoran cepat saji. Restoran cepat saji merupakan tempat pelayanan makanan dan minuman kepada konsumen secara cepat. Berdasarkan data Badan Pusat Statistik (BPS), bahwa jumlah usaha penyediaan makanan minuman tercatat 4,85 juta usaha atau mengalami peningkatan sebesar 21,13% pada tahun 2023 dibandingkan tahun 2016 yang hanya mencapai 4,01 juta usaha [1]. Serta hasil data BPS tentang statistik penyediaan makanan dan minuman diambil pada tahun 2019, 2020, 2021, dan 2023 telah menunjukkan peningkatan jumlah restoran pada setiap tahunnya di Indonesia [2].

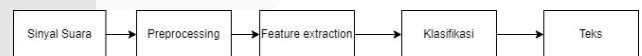
Dengan banyak persaingan bisnis restoran cepat saji sehingga perusahaan menyediakan jasa layanan yaitu layanan *drive thru*. Sistem layanan *drive thru* saat ini menghadapi

tantangan dalam efisiensi dan keakuratan dalam pemesanan *drive thru*. Tujuan penelitian ini adalah membuat rancangan dan implementasi sistem *voice recognition* pada layanan *drive thru* dengan bahasa pemrograman Python. Teknologi ini dapat mengonversikan gelombang suara menjadi bentuk teks. Hasil teks tersebut akan dikirimkan ke pramusaji yang berisi pesanan sehingga dapat meminimalisir kesalahan pesanan produk makanan dan minuman dari konsumen.

II. KAJIAN TEORI

A. *Voice recognition*

Voice recognition merupakan suatu sistem yang mengonversi kalimat suara menjadi format digital dalam berupa kode-kode digital [3]. Teknologi *voice recognition* yang suara yang dapat menjadi landasan hubungan manusia dengan komputer modern sebagai komunikasi antara pengguna dan perangkat melalui perintah yang diucapkan. Proses *voice recognition* untuk mengubah sinyal suara menjadi bentuk teks yaitu preprocessing, feature extraction, dan klasifikasi [4].



GAMBAR 1

(PROSES *VOICE RECOGNITION*)

Gambar diatas, masukkan sinyal suara ke preprocessing untuk mempersiapkan dan mengolah data agar data siap dipakai sebelum ke tahapan *feature extraction*. *Feature extraction* yaitu tahapan mengidentifikasi dan mengekstrak dalam bentuk data yang mentah seperti suara asli dengan bertujuan meningkatkan akurasi [5]. Lalu diklasifikasikan berdasarkan kelas untuk melakukan pengenalan pola-pola pada setiap kelas dan diubah menjadi bentuk teks [6].

B. Short Time Fourier Transform

Short Time Fourier Transform (STFT) adalah sebuah metode lanjutan dari fourier transform yang mengonversi sinyal dari domain waktu menjadi ke domain frekuensi kemudian membagi ucapan menjadi beberapa *frame* [7]. Metode STFT memiliki kelebihan yaitu dapat menampilkan perubahan frekuensi yang berdasarkan satuan waktu [8].

STFT digunakan sebagai ekstraksi fitur yang dilakukan menganalisis spektrum dan mengetahui karakteristik frekuensi sinyal suara. Ada beberapa proses dalam STFT yaitu *framing*, *windowing*, *transformation*, dan *concatenation* [9].

Persamaan (1) merupakan perhitungan STFT berdasarkan waktu dan frekuensi dibawah ini.

$$X(m, \omega) = \sum_{n=-\infty}^{\infty} x(n)w(n-m)e^{-j\omega n} \quad (1)$$

Keterangan :

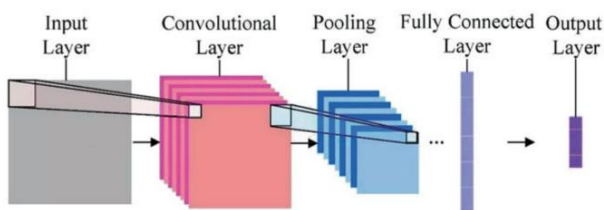
$X(m, \omega)$: Representasi spektrum sinyal pada waktu m dan frekuensi ω

$x(n)$: Sinyal asli dalam domain waktu

$w(n)$: Fungsi window yang berupa fungsi Hamming, Hanning, atau Gaussian

C. Convolutional Neural Network

Convolutional Neural Network (CNN) merupakan suatu model kategori *deep learning* untuk mengklasifikasikan, mengidentifikasi, dan mengenali pola dalam citra yang telah diekstraksi fitur dari sinyal audio [10]. Model ini menggunakan metode lapisan konvolusi dibangun dengan library TensorFlow. Komponen yang diperlukan yaitu data masukan, filter, dan peta fitur. Pada gambar 2 menunjukkan lapisan model CNN yang terdiri dari tiga lapisan yaitu *convolutional layer*, *pooling layer*, dan *fully connected layer*[11].



GAMBAR 2
(STRUKTUR CNN)

Pada proses konvolusi, filter bergerak melintasi area gambar agar dapat memperoleh komputasi *dot product*. Hasil operasi *dot product* disebut peta fitur, setiap filter menghasilkan satu peta fitur dan jumlah filter yang digunakan untuk menentukan jumlah fitur [12]. Lapisan *pooling* bertugas mengurangi ukuran representasi data kompleksitas model. Lapisan ini memiliki dua jenis pooling yaitu *maxpooling* dan *average pooling*[13]. *Maxpooling* adalah filter yang memiliki piksel dengan nilai maksimum dari setiap patch pada peta fitur. *Average pooling* yaitu filter yang menghitung rata-rata untuk setiap patch. Kemudian lapisan *fully connected* berfungsi untuk mengklasifikasikan berdasarkan fitur-fitur yang telah diekstraksi sebelumnya. Hasil keluarannya yaitu probabilitas dalam kelas yang menunjukkan seberapa yakin model pada setiap kelas [14].

D. Confusion Matrix

Confusion matrix adalah suatu alat yang memudahkan perhitungan metrik evaluasi seperti akurasi, presisi, recall, dan F1 score dalam *machine learning*. Matrik ini memiliki empat kondisi utama yang menilai kinerja model dalam klasifikasi yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Empat komponen tersebut berperan penting untuk mencari nilai akurasi, presisi, recall, dan F1-score [15].

Perhitungan metrik evaluasi meliputi sebagai berikut:

$$Akurasi = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (2)$$

$$Presisi = \frac{TP}{(TP+FP)} \quad (3)$$

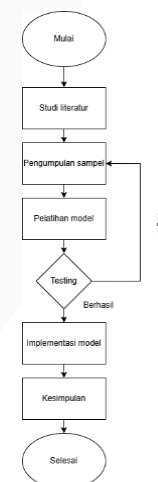
$$Recall = \frac{TP}{(TP+FN)} \quad (4)$$

$$F1-Score = \frac{2TP}{(2TP+FP+FN)} \quad (5)$$

Berikut penjelasan persamaan metrik evaluasi yang diperoleh dari confusion matrix. Persamaan (2) adalah Akurasi, ukuran seberapa sering prediksi yang telah dibuat model secara keseluruhan data. Persamaan (3) adalah presisi, ukuran seberapa akurat model untuk memprediksi kelas tertentu dengan benar. Persamaan (4) adalah recall, ukuran untuk mengetahui berapa banyak kasus positif yang sebenarnya dalam memprediksi dengan benar. Kemudian persamaan (5) adalah F1 score, metrik yang digunakan untuk mengukur dan memberikan penilaian kinerja model yang seimbang antara presisi dan recall.

III. METODE

A. Diagram Alur Penelitian

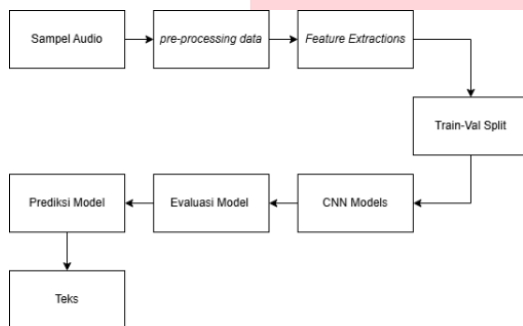


GAMBAR 3
(DIAGRAM ALUR PENELITIAN)

Pada gambar diatas merupakan diagram alur penelitian yang dimulai dari studi literatur, pengumpulan sampel audio, dan menjalankan pelatihan model. Jika testing gagal maka perlu melakukan pengumpulan sampel audio untuk pelatihan model yang baru. Apabila testing berhasil, melakukan pengimplementasi model dan membuat kesimpulan dari penelitian.

B. Pelatihan Model *Voice Recognition*

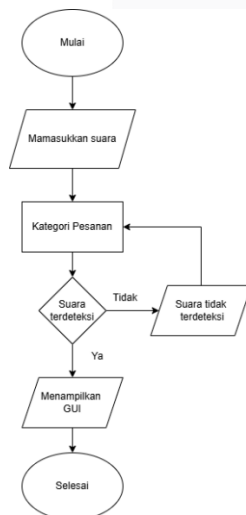
Pelatihan model *voice recognition* menunjukkan urutan tahapan melatih model *voice recognition* secara terstruktur. Tahapan awal diawali *preprocessing* yaitu mempersiapkan dan mengolah data sebelum melakukan pelatihan model. Selanjutnya *feature extraction*, mengurangi dimensi data dan mengekstrak fitur – fitur sampel data. Kemudian melakukan pembagian set pelatihan dan set test untuk model dapat mengenali pola – pola dan generalisasi data yang baru. Model dibangun dengan model CNN untuk pengukuran metrik evaluasinya. Setelah evaluasi model, pengujian model yang telah dilatih dengan data yang pernah dilatih dan data yang baru. Hal ini menguji seberapa baiknya model untuk memprediksi kategori dengan data aktualnya. Hasil pelatihan model tersebut disimpan dalam format h5 dan diimplementasikan secara nyata pada sistem.



GAMBAR 4
(MELATIH MODEL *VOICE RECOGNITION*)

C. Prosedur Kerja Sistem *Voice Recognition*

Prosedur kerja sistem *voice recognition* merupakan rangkaian cara kerja menggunakan sistem *voice recognition* pada layanan drive thru restoran cepat saji secara terstruktur yang dapat dilihat pada gambar dibawah ini.



GAMBAR 5
(DESAIN BLOK SISTEM *VOICE RECOGNITION*)

Gambar 5 menunjukkan hasil model yang telah latih sebelumnya dan integrasikan ke sistem layanan drive thru. Rancangan sistem ini dimulai memasukkan suara melalui mikrofon, pengguna memilih pesanan yang ada dimenu berdasarkan kategori pesanan. Jika suara tidak terdeteksi maka sistem mengharuskan pengguna untuk memilih

pesanan ulang. Apabila suara terdeteksi maka suara menjadi bentuk teks yang akan ditampilkan melalui GUI sesuai dengan pesanan pengguna.

D. Sumber Data

Sumber data sampel audio yang berdasarkan jenis kelamin yaitu laki-laki dan perempuan. Sampel audio diperoleh dari 30 responden yaitu 19 orang laki-laki dan 11 orang Perempuan. Rekaman sumber data diambil orang-orang di sekitar kampus dan rumah. perekaman sampel audio menggunakan aplikasi Matlab yang terkoneksi dengan mikrofon. Sampel frekuensi yang digunakan yaitu 16 kHz dan kanalnya mono. Lokasi dilakukan pada lingkungan yang kedap suara. Hal ini mengurangi kebisingan sampel suara dan menjaga kualitas suara saat pelatihan model. Tabel 1 merupakan jumlah data dari setiap kelas yang telah diaugmentasi data.

TABEL 1
(KATEGORI KELAS)

No	Kelas	Jumlah
1.	<i>Soup</i>	270
2.	<i>Cake</i>	270
3.	<i>Chicken</i>	270
4.	<i>Burger</i>	270
5.	<i>Rice</i>	270
6.	<i>Juice</i>	270
7.	<i>Coffee</i>	270
8.	<i>Tea</i>	270
9.	<i>Water</i>	270
10.	<i>Milk</i>	270
11.	<i>Small</i>	270
12.	<i>Medium</i>	270
13.	<i>Large</i>	270
Total Data		3.510

E. Metode Short Time Fourier Transform

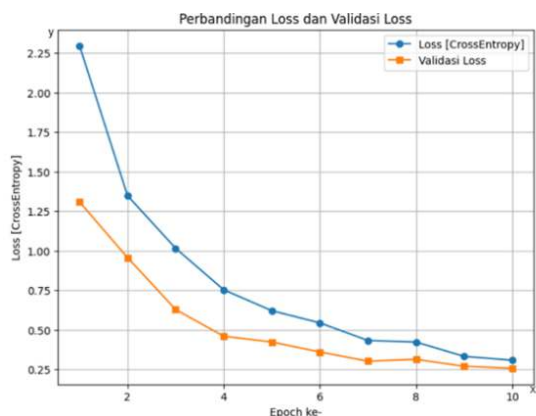
Metode Short Time Fourier Transform digunakan dalam penelitian untuk merepresentasikan bentuk waveform dari sinyal domain waktu menjadi sinyal domain frekuensi- waktu untuk menghitung metode STFT. Hasil metode STFT direpresentasikan menjadi gambar 2D yaitu spektrogram. Spektrogram terdiri dari dua intensitas warna yaitu warna terang dan gelap. Warna terang menandakan frekuensi tersebut memiliki amplitudo yang rendah. Warna gelap menandakan frekuensi memiliki amplitudo yang tinggi. Metode ini membagi sinyal audio menjadi beberapa segemen dalam frame dengan mempertahankan beberapa informasi waktu. Hasil spektrogram akan dilanjutkan untuk menjalankan model CNN. CNN akan menjalankan lapisan *konvolusi*, lapisan *pooling*, dan lapisan *fully connected* agar dapat mengekstraksi warna yang diambil dari spektrogram. Hasil ekstraksi tersebut akan diperoleh metrik evaluasi dari confusion matrix seperti akurasi, presisi, recall, dan F1-score.

IV. HASIL DAN PEMBAHASAN

A. Hasil Pengujian Nilai Akurasi dan Loss

Hasil pengujian yang dilakukan untuk melatih model selama 10 epoch. Epoch bagi model yaitu dapat memperbarui bobot dan bias berdasarkan kesalahan yang dihitung setelah memproses seluruh data dalam *machine learning*. Setiap epoch memiliki tingkatan nilai akurasi dan nilai *loss* yang

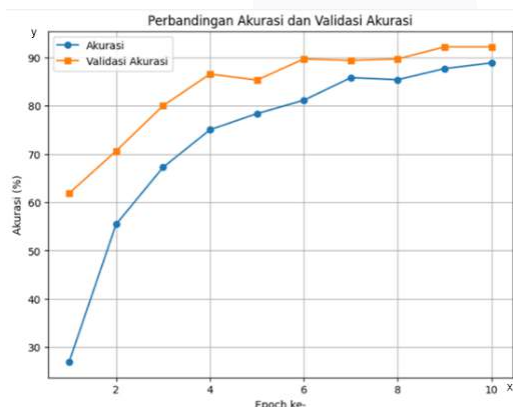
berbeda sehingga grafik yang dihasilkan naik atau turun. Pengukuran nilai akurasi dan *loss* disajikan pada Gambar 6 dan Gambar 7 sebagai berikut:



GAMBAR 6

(GRAFIK *TRAINING LOSS* DAN *VALIDASI LOSS*)

Gambar 6 merupakan hasil grafik pengujian melatih model dari *loss* pada pelatihan dan validasi. Berdasarkan grafik tersebut perbandingan kedua linear bahwa *training loss* dan validasi *loss* mengalami penurunan. Pada sumbu x adalah jumlah epoch hingga 10 epoch dan sumbu y adalah nilai *loss*. Penurunan dari kedua grafik merupakan indikasi yang positif terhadap kinerja model karena model dapat belajar baik dari data pelatihan dan mampu generalisasi terhadap data validasi.



GAMBAR 7

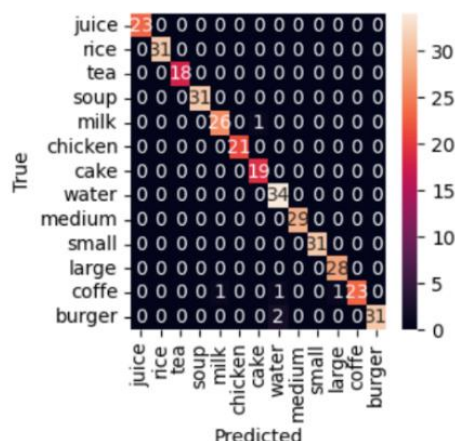
(GRAFIK *TRAINING AKURASI* DAN *VALIDASI AKURASI*)

Gambar 7 merupakan hasil grafik pengujian melatih model dari akurasi pada pelatihan dan validasi. Diketahui grafik tersebut menunjukkan kenaikan. Jika *training* akurasi yang meningkat maka model dapat mengenali dan memproses pola - pola selama pelatihan. Apabila validasi akurasi yang naik menandakan bahwa model mampu generalisasi terhadap data yang baru.

Berdasarkan hasil kedua grafik gambar 6 dan 7 menunjukkan hasil positif yang signifikan dikarenakan grafik *loss* mengalami penurunan seiring bertambahnya setiap epoch. Penurunan mengindikasikan bahwa seberapa jauhnya model dari nilai sebenarnya. Sedangkan grafik akurasi mengalami peningkatan seiring bertambahnya setiap epoch. Peningkatan ini menunjukkan seberapa baik model membuat

prediksi yang benar. Kedua grafik tersebut model dapat memprediksi dengan data dan generalisasi data yang baru.

B. Perhitungan Confusion Matrix



GAMBAR 8

(CONFUSION MATRIX)

Dapat dilihat Gambar 8 merupakan *confusion matrix* yang terdiri dari 13 kelas yaitu *rice*, *soup*, *chicken*, *cake*, *burger*, *milk*, *water*, *tea*, *coffee*, *juice*, *medium*, *small*, dan *large*. Komponen tersebut digunakan untuk mengukur metrik evaluasi seperti akurasi, presisi, recall, dan F1-score dari seluruh kelas. Setiap kelas berisi angka probabilitas dari TP, FP, FN, dan TN pada tabel 2.

TABEL 2

(PERHITUNGAN DARI CONFUSION MATRIX)

Kategori	Kelas	TP	FP	FN	TN
A	<i>Soup</i>	31	0	0	320
B	<i>Cake</i>	19	0	1	331
C	<i>Chicken</i>	21	0	0	330
D	<i>Burger</i>	31	2	0	318
E	<i>Rice</i>	31	0	0	320
F	<i>Juice</i>	23	0	0	328
G	<i>Coffee</i>	23	3	0	325
H	<i>Tea</i>	18	0	0	333
I	<i>Water</i>	34	0	3	314
J	<i>Milk</i>	26	1	1	323
K	<i>Small</i>	31	0	0	320
L	<i>Medium</i>	29	0	0	322
M	<i>Large</i>	28	0	1	322
Total		345	6	6	4206

Hasil pengukuran yang didapatkan dari TP, FP, FN, dan TN akan memperoleh nilai metrik evaluasinya. Nilai metrik evaluasi dapat dilihat pada tabel 3.

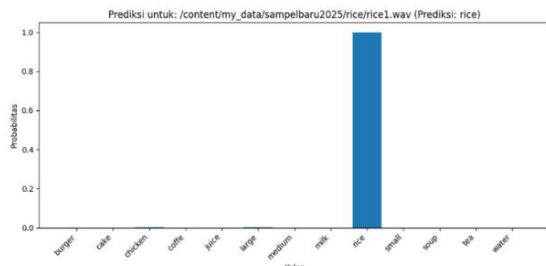
TABEL 3

(NILAI METRIK EVALUASI)

No	Paramater	simulasi	Perhitungan manual
1.	Accuracy	98 %	99,737 %
2.	Precision	98 %	98,3613 %
3.	Recall	98 %	98,4415 %
4.	F1-Score	98 %	0,982906

Pada tabel 3 merupakan tabel perbandingan nilai simulasi dan perhitungan manual dari metrik evaluasi dalam penelitian. Pengukuran parameter yaitu akurasi, presisi, recall, dan F1-score. Tujuan perbandingan ini adalah dapat mengetahui seberapa akurat penggunaan simulasi dibandingkan perhitungan manual. Nilai seluruh parameternya telah mencapai 98 % dari nilai simulasi dan perhitungan manual. Hasil pengukuran dari simulasi menandakan adanya pendekatan dengan nilai hasil perhitungan manual sehingga model dapat mempelajari dan memproses pola - pola yang baik dari pengujian penelitian.

C. Hasil Pengujian Model Selama Pelatihan



GAMBAR 9
(DIAGRAM PENGUJIAN MODEL VOICE RECOGNITION)

Pada gambar diatas merupakan hasil pengujian model *voice recognition* yang telah dilatih sebelumnya. Hal ini bertujuan untuk mengetahui model dapat memprediksi data sebelumnya dan data yang baru. Pengujian *rice* ditunjukkan kelas *rice* paling tinggi dibandingkan yang lain bahwa data kelas ini sesuai dengan data aktual. Hal ini menandakan bahwa model dapat memprediksi kelas dengan baik berdasarkan data yang telah diinputkan. Model yang telah diuji dapat disimpan dalam format h5 dan siap diimplementasikan.

D. Hasil Pengujian Tingkat Akurasi pada Setiap Kelas

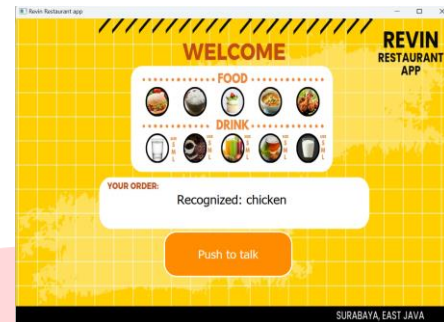
TABEL 4
(TINGKAT AKURASI SETIAP KELAS)

No	Kelas	Tingkat Akurasi
1.	<i>Soup</i>	85 %
2.	<i>Cake</i>	71 %
3.	<i>Chicken</i>	86 %
4.	<i>Burger</i>	89 %
5.	<i>Rice</i>	69 %
6.	<i>Juice</i>	84 %
7.	<i>Coffee</i>	66 %
8.	<i>Tea</i>	93 %
9.	<i>Water</i>	54 %
10.	<i>Milk</i>	83 %
11.	<i>Small</i>	68 %
12.	<i>Medium</i>	94 %
13.	<i>Large</i>	80 %

Pada tabel 4 menunjukkan hasil pengujian tingkat akurasi setiap kelas pada model yang telah dilatih dalam pengujian model dalam sistem. Pengujian sistem deteksi ini dilakukan sebanyak 100 kali percobaan dan telah integrasikan ke sistem

melalui Visual Studio Code. Hasil menunjukkan bahwa tea sebesar 93 % dan medium sebesar 94 % memiliki tingkat akurasi tertinggi. Kemudian kelas water yang paling rendah sebesar 54 %. Hal ini menunjukkan sistem dapat memprediksi kelas tea dan medium dengan baik

E. Hasil Pengujian Model dalam Sistem



GAMBAR 10
(GUI PENGGUNA)

Dapat dilihat Gambar 10 adalah hasil pengujian yang disimpan dalam format h5 telah diintegrasikan ke sistem *voice recognition* dan menampilkan GUI kepada pengguna. Pengguna hanya menekan pesanan dan memilih pesanan sesuai dengan kategori sebelumnya yang telah dilatih. Pesanan akan muncul dalam bentuk teks seperti pada gambar diatas. Library digunakan salah satunya yaitu Tensorflow, TensorFlow berperan untuk membangun model yang telah disimpan dan diaplikasikan dengan PyQt5 untuk menampilkan GUI.



GAMBAR 11
(GUI SUARA TIDAK TERDETEKSI)

Gambar 11 menunjukkan hasil suara yang tidak terdeteksi oleh sistem. Apabila pengguna tidak mengucapkan pesanan maka sistem tidak dapat mendeteksi adanya suara yang dipesan melalui mikrofon. Hal ini berfungsi untuk mengurangi suara yang tidak diinginkan dan meningkatkan penerimaan kualitas suara yang masuk ke sistem.

V. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan dapat disimpulkan bahwa perancangan dan implementasi sistem *voice recognition* berbasis AI pada layanan *drive thru* restoran cepat saji telah berhasil. Pengujian melatih model mendapat nilai metrik evaluasi yang baik.

Penelitian menunjukkan hasil pengujian dalam model *voice recognition* dari akurasi sebesar 98 %, presisi sebesar

98 %, recall sebesar 98%, dan F1-score 0,98. Seiring bertambahnya epoch, grafik akurasi mengalami peningkatan dan *loss* mengalami penurunan. Hal ini menandakan bahwa model dapat memprediksi dari beberapa kelas. Hasil perhitungan simulasi memiliki nilai yang berdekatan dengan hasil perhitungan manual.

Hasil implementasi kelas *tea* dan *medium* tingkat akurasi terbesar. Kemudian tingkat akurasi yang terkecil yaitu kelas *water*. Ada beberapa faktor yang mempengaruhi hasil tingkat akurasi pada setiap kelas memiliki nilai yang berbeda – beda. Faktor utama dalam sistem *voice recognition* yaitu sampel suara yang bervariasi seperti kecepatan suara, intonasi suara, dan aksent. Juga penyesuaian lingkungan yang *noise* agar model memiliki tingkat akurasi yang tinggi pada setiap kelas.

REFERENSI

- [1] Badan Pusat Statistik, *Buku Statistik Penyediaan Makanan dan Minuman 2023*. Jakarta : Badan Pusat Statistik, 2024.
- [2] Badan Pusat Statistik, *Buku Statistik Penyediaan Makanan dan Minuman 2020*. Jakarta : Badan Pusat Statistik, 2021.
- [3] A. Devitra dan R. Purbaningtyas, "Prototype Smart Home System Menggunakan Voice Control Pada Perangkat Iot," *Just IT J. Sist. Informasi, Teknol. Inf. dan Komput.*, vol. 13, no. 1, hal. 53–59, 2022.
- [4] R. M. R. Clinton dan S. Sengkey, "Purwarupa Sistem Daftar Pelanggaran Lalulintas Berbasis Mini-Komputer Raspberry Pi," *J. Tek. Elektro dan Komput. Vol.8*, vol. 8, no. 3, hal. 181–192, 2019.
- [5] D. Al-Fraihat, Y. Sharrab, F. Alzyoud, A. Qahmash, M. Tarawneh, dan A. Maaita, "Speech Recognition Utilizing Deep Learning: A Systematic Review of the Latest Developments," *Human-centric Comput. Inf. Sci.*, vol. 14, no. March, 2024, doi: 10.22967/HGIS.2024.14.015.
- [6] F. Adnan, I. Amelia, dan S. 'Umar Shiddiq, "Implementasi Voice Recognition Berbasis Machine Learning," *Implementasi Voice Recognit. Berbas. Mach. Learn.*, vol. 11, no. 1, hal. 24–29, 2022.
- [7] H. R. Paleva dan B. H. Prasetyo, "Penerapan Short Time Fourier Transform pada MFCC untuk Sistem Pengenalan Ucapan Tingkat Stres," vol. 1, no. 1, hal. 1–10, 2024.
- [8] A. H. Prawironegoro dan B. H. Prasetyo, "PENERAPAN SHORT TIME FOURIER TRANSFORM PADA PEMBICARA," *Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 1, 2017.
- [9] S. F. Salsabillah *et al.*, "Short Time Fourier Transform (STFT) Sebagai Feature Extraction Deteksi Kerusakan Inner Race Bearing Motor Induksi Secara Realtime Menggunakan Sinyal Suara," *J. Tek. Elektro*, vol. 06, no. 02, hal. 20–26, 2023.
- [10] IBM, "What is Convolutional Neural Network," IBM.com. Diakses: 5 Januari 2025. [Daring]. Tersedia pada: <https://www.ibm.com/id-id/topics/convolutional-neural-networks>
- [11] R. D. Nurfitra dan G. Ariyanto, "Implementasi Deep Learning berbasis Tensorflow untuk Pengenalan Sidik Jari," *Emit. J. Tek. Elektro*, vol. 18, no. 1, hal. 22–27, 2018, doi: 10.23917/emit.v18i01.6236.
- [12] BPJIIID, "Convolutional Neural Network (CNN): Teknologi Kecerdasan Buatan untuk Pengenalan Gambar," bpjiid.uma.ac.id. Diakses: 5 Januari 2025. [Daring]. Tersedia pada: <https://bpjiid.uma.ac.id/2024/09/07/convolutional-neural-networks-cnn-teknologi-kecerdasan-buatan-untuk-pengenalan-gambar/>
- [13] A. ANHAR dan R. A. PUTRA, "Perancangan dan Implementasi Self-Checkout System pada Toko Ritel menggunakan Convolutional Neural Network (CNN)," *ELKOMIKA J. Tek. Energi Elektr. Tek. Telekomun. Tek. Elektron.*, vol. 11, no. 2, hal. 466, 2023, doi: 10.26760/elkomika.v11i2.466.
- [14] Sufriana, "Penerapan CNN (Convolutional Neural Network) dalam Mengidentifikasi Varietas Buah Berbasis Citra Digital," purwadhika.com. Diakses: 5 Januari 2025. [Daring]. Tersedia pada: <https://purwadhika.com/blog/penerapan-cnn-convolutional-neural-network-dalam-mengidentifikasi-varietas-buah-berbasis-citra-digital>
- [15] D. Normawati dan S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, hal. 697–711, 2021.