

SYSTEMIC: Information System and Informatics Journal

ISSN: 2460-8092, 2548-6551 (e)

Vol 7 No 2 – Desember 2021

Implementasi Model Regresi Logistik dalam Klasifikasi Kebutuhan Ruang ICU Terhadap Pasien Positif COVID-19Baharudin Pratama¹, Natalia Damastuti²^{1,2}) Universitas Narotama, Surabayabaharudinp21@gmail.com¹, natalia.damastuti@narotama.ac.id²**Kata Kunci***COVID-19, ICU, Data Science, Regresi Logistik***Abstrak**

Coronavirus Disease 19 (COVID-19) merupakan jenis penyakit yang disebabkan oleh suatu virus bernama SARS-CoV-2. Statistik menyatakan bahwa 1 dari 5 pasien COVID-19 menjalani perawatan di rumah sakit, dan 1 dari 10 diantaranya membutuhkan penanganan di ruang ICU (Intensive Care Unit). Pada penelitian ini dilakukan klasifikasi kebutuhan ruang ICU dilakukan pada pasien COVID-19 berdasarkan penyakit penyerta dan kondisi tertentu menggunakan model regresi logistik. Regresi logistik digunakan dengan pertimbangan data dan variabel penelitian memiliki skala data kategori. Data dibagi menjadi dua yaitu data training dan data testing dengan rasio perbandingan 80%:20%. Tujuan penelitian adalah mendapatkan nilai akurasi dari klasifikasi yang dilakukan. Hasil dari confusion matrix dari proses klasifikasi menunjukkan true positive sebesar 4,92%, true negative sebesar 82,37%, false positive sebesar 4,63%, dan false negative sebesar 8,09%. Berdasarkan hasil tersebut diperoleh nilai akurasi 87,29%.

Keywords*COVID-19, ICU, Data Science, Logistic Regression***Abstract**

Coronavirus Disease 19 (COVID-19) is a type of disease caused by a virus called SARS-CoV-2. Statistics state that 1 out of 5 COVID-19 patients undergo hospital treatment, and 1 out of 10 of them require treatment in the ICU (Intensive Care Unit). In this study, the classification of the need for ICU space was carried out in COVID-19 patients based on accompanying diseases and specific conditions using a logistic regression model. Logistic regression is used with consideration of the data and the research variable has a categorical data scale. The data is divided into two, namely training data and testing data with a comparison ratio of 80%:20%. The purpose of the research is to get the accuracy value of the classification done. The results of the confusion matrix from the classification process show that true positive is 4.92%, true negative is 82.37%, false positive is 4.63%, and false negative is 8.09%. Based on these results, an accuracy value of 87.29% was obtained.

1. Pendahuluan

COVID-19 disebabkan oleh infeksi virus SARS-CoV-2. Gejala yang terjadi ketika terinfeksi virus tersebut terdiri dari demam, batuk-batuk, hidung tersumbat, kelelahan, serta masalah lain terkait infeksi saluran pernapasan atas [1]. Virus tersebut menginfeksi organ sistem pernapasan dan dapat menyebar ke organ lain seperti, jantung, pembuluh darah, ginjal, usus, dan otak [2]. Laporan infeksi pertama kali diketahui berasal dari Pasar *Seafood* Huanan di Kota Wuhan, yang merupakan Ibu Kota Provinsi Hubei, Republik Rakyat Tiongkok. Dianggap sebagai ancaman kesehatan publik pada Desember 2019, kemudian pada Januari 2020 *World Health Organization*

(WHO) menyatakan kondisi darurat kesehatan publik [3]. Selanjutnya WHO menyatakan COVID-19 sebagai pandemik pada Maret 2020 setelah menyebar ke lebih dari seratus negara di dunia dan menyebabkan ribuan kasus kematian [4].

Infeksi virus SARS-CoV-2 yang menyerang sistem pernapasan dan dapat menyebar ke organ lain yang mengancam keselamatan hidup pasien terinfeksi, hal tersebut menyebabkan pasien membutuhkan perawatan yang intensif dan bantuan alat pernapasan mekanik. Statistik menyatakan bahwa 1 dari 5 pasien COVID-19 menjalani perawatan di rumah sakit, dan 1 dari 10 diantaranya membutuhkan penanganan di ruang ICU (*Intensive Care Unit*) [5]. Pertimbangan kebutuhan untuk perawatan ruang ICU pasien

COVID-19 diklasifikasikan berdasarkan kriteria-kriteria yang menjadi pertimbangan seperti penyakit penyerta (komorbid) atau kondisi tertentu seperti mengandung.

Berikut keterkaitan antara COVID-19 dengan penyakit penyerta atau kondisi tertentu pada pasien:

- a. Peradangan paru-paru (*pneumonia*)
Peradangan paru-paru biasa terjadi dalam dua hingga tiga minggu setelah infeksi, ditandai dengan berkurangnya kadar oksigen, gas darah tidak normal, dan perubahan yang bisa terlihat pada dada ketika dilakukan X-ray. Infeksi berpeluang memicu penyakit berat serta 75% pasien memiliki risiko sesak napas berdasarkan komputasi tomografi pada admisi [1]. *Pneumonia* sebagai komorbid COVID-19 menunjukkan tingkat kematian hingga 30-35% pada pasien yang membutuhkan alat bantu pernafasan mekanik [6].
- b. Diabetes
Pasien COVID-19 penderita diabetes mellitus meningkatkan risiko admisi ruang ICU dan keselamatan hidup terinfeksi COVID-19 [7].
- c. COPD (*Chronic Obstructive Pulmonary Disease*)
COPD, dalam Bahasa Indonesia PPOK (Penyakit Paru Obstruktif Kronik) merupakan salah satu infeksi saluran pernapasan. COPD menjadi salah satu faktor risiko infeksi COVID-19 menjadi lebih parah [8]. Beberapa studi melaporkan COPD menambah risiko admisi ICU dan gagal napas [9].
- d. Asma
Seperti virus lain, SARS-CoV-2 juga dapat memicu eksaserbasi (serangan) asma. Hal tersebut meningkatkan risiko keselamatan hidup pasien COVID-19 yang menderita asma, yang mana menjadikan asma sebagai salah satu parameter pada data pasien [10].
- e. Imunosupresi
Merupakan kondisi penekanan sistem imun. Kondisi tersebut pada dasarnya menurunkan imunitas (daya tahan) tubuh sehingga meningkatkan risiko infeksi oleh bakteri, virus, atau jamur, termasuk meningkatnya risiko infeksi SARS-CoV-2 [11].
- f. Hipertensi
Hipertensi memiliki keterkaitan terhadap bertambahnya risiko kematian pada pasien COVID-19, sindrom pernapasan akut, kebutuhan rawat ruang ICU, dan perkembangan penyakit pada pasien COVID-19 [12].
- g. Kardiovaskular
Kardiovaskular merupakan komorbiditas (penyakit penyerta) pada terinfeksi COVID-19 meningkatkan risiko keselamatan hidup pasien [13].
- h. Obesitas
Obesitas memiliki keterkaitan dengan COVID-

19 menyebabkan kondisi lebih buruk pada pasien. Penderita obesitas memiliki risiko infeksi sistem saluran pernapasan bawah yang mana berpotensi menambah infeksi lain [14].

- i. Gagal ginjal
Infeksi COVID-19 pada seseorang yang mengalami kerusakan ginjal atau penderita gagal ginjal dapat meningkatkan risiko keselamatan hidup. Hal tersebut juga terjadi pada pasien yang memiliki permasalahan pada ginjal serta pasien transplantasi [3].
- j. Merokok
Perokok dengan konsumsi rokok tradisional maupun rokok elektrik memiliki risiko lebih tinggi terinfeksi COVID-19 dibandingkan non-perokok. Temuan bahwa aktifitas merokok berkaitan dengan COVID-19 selaras dengan efek rokok yaitu mengganggu imunitas paru-paru [15].
- k. Kehamilan
Wanita yang terinfeksi COVID-19 dengan kondisi mengandung (hamil) lebih dimungkinkan untuk mendapatkan penanganan rumah sakit dan admisi ICU untuk mendapatkan bantuan alat pernapasan dibandingkan wanita yang tidak mengandung [16].

Kondisi yang menyertai infeksi virus tersebut mampu meningkatkan ancaman terhadap keselamatan hidup pasien. Rekomendasi melakukan *social distancing*, menggunakan masker, dan sering melakukan cuci tangan sebagai upaya pencegahan yang dapat dilakukan guna mengurangi risiko penyebaran infeksi virus SARS-CoV-2.

Keterkaitan kondisi-kondisi tersebut dengan infeksi virus SARS-CoV-2 menjadi pertimbangan untuk admisi ICU terhadap pasien COVID-19. Klasifikasi kebutuhan ruang ICU terhadap pasien COVID-19 dapat dilakukan dalam model regresi logistik dengan mempertimbangkan penelitian lain yang telah dilakukan dengan implementasi regresi logistik pada *dataset* dalam bidang kesehatan.

Tinjauan penelitian oleh Siti Masrofatul Azizah dan Novita Eka Chandra (2017) dengan judul Model Regresi Logistik pada Faktor-Faktor yang Mempengaruhi Imunisasi Lengkap Balita dengan hasil tingkat akurasi sebesar 84,8% [17].

Tinjauan penelitian selanjutnya oleh Syed Hasan Adil dkk. (2018) dengan judul *Liver Patient Classification using Logistic Regression* dengan sumber data *open source dataset* yang dipublikasikan di UCI *machine learning* dengan nama *Indian Liver Patient Records*. Klasifikasi telah dilakukan pada *dataset* pada penelitian-penelitian sebelumnya menggunakan teknik yang berbeda. Analisis menghasilkan nilai akurasi sebesar 74%. Nilai akurasi tersebut lebih tinggi dibandingkan dengan teknik klasifikasi lain pada *dataset* yang sama [18].

Tinjauan penelitian selanjutnya oleh Montu Saw dkk. (2020) dengan judul *Estimation of Prediction for Getting Heart Disease Using Logistic Regression Model of Machine Learning*. Penelitian tersebut bertujuan meningkatkan tingkat akurasi pada prediksi risiko seseorang terkena penyakit jantung dalam rentang waktu 10 tahun kedepan. Hasil analisis menunjukkan model *logistic regression* mampu melakukan klasifikasi dengan hasil nilai akurasi sebesar 87%, lebih tinggi sekitar 0,7% dari penelitian sebelumnya [19].

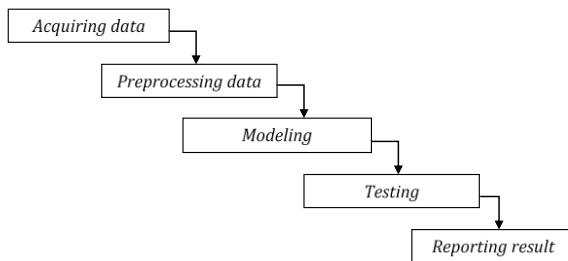
Penelitian ini dilakukan dengan tujuan untuk mengetahui tingkat akurasi pada klasifikasi kebutuhan ruang ICU untuk pasien COVID-19 menggunakan regresi logistik.

2. Dasar Teori

2.1 Data Science

Data science didefinisikan sebagai proses ekstraksi informasi yang berarti dari sebuah *dataset*, dengan metode analisis yang berasal dari beberapa bidang ilmu seperti ilmu komputer, matematika, statistik, *data warehouse*, dan proses distribusi. Istilah *data science* pada pencarian di internet dikaitkan dengan *data analytics*, *predictive analytics*, *knowledge discovery from data*, *machine learning*, dan *data mining*. Istilah-istilah tersebut dimaknai secara luas sebagai *data science* [20].

Proses analisis data menghasilkan model untuk penyelesaian masalah, *modeling data* hanya dibangun setelah dibaca dan dilakukan *data preprocessing* sesuai kebutuhan dengan proses yang memiliki tahapan dasar pada Gambar x.



Gambar 1. Tahapan Dasar Data Science

2.2 Sumber Data

Dataset yang digunakan yaitu *COVID-19 patient pre-condition dataset* yang diakses dan diunduh dari *open source* data Kaggle. *Dataset* tersebut memiliki lisensi *Public Domain* sehingga bisa digunakan oleh publik secara leluasa termasuk melakukan penelitian. Berikut link untuk mengakses data yang digunakan <https://www.kaggle.com/datasets/tanmoiyx/covid19-patient-precondition-dataset>.

2.3 Variabel Penelitian

Daftar variabel yang digunakan dalam

melakukan penelitian berupa fitur data mengenai kondisi pasien COVID-19 pada *dataset*.

Tabel 1. Variabel Penelitian

Fitur	Skala Data	Kategori
<i>sex</i>	nominal	(1) <i>female</i> (2) <i>male</i>
<i>intubed</i>	nominal	(1) <i>yes</i> (2) <i>no</i>
<i>pneumonia</i>	nominal	(1) <i>yes</i> (2) <i>no</i>
<i>age</i>	rasio	-
<i>pregnancy</i>	nominal	(1) <i>yes</i> (2) <i>no</i>
<i>diabetes</i>	nominal	(1) <i>yes</i> (2) <i>no</i>
<i>copd</i>	nominal	(1) <i>yes</i> (2) <i>no</i>
<i>asthma</i>	nominal	(1) <i>yes</i> (2) <i>no</i>
<i>inmsupr</i>	nominal	(1) <i>yes</i> (2) <i>no</i>
<i>hypertension</i>	nominal	(1) <i>yes</i> (2) <i>no</i>
<i>other disease</i>	nominal	(1) <i>yes</i> (2) <i>no</i>
<i>cardiovaskular</i>	nominal	(1) <i>yes</i> (2) <i>no</i>
<i>obesity</i>	nominal	(1) <i>yes</i> (2) <i>no</i>
<i>renal chronic</i>	nominal	(1) <i>yes</i> (2) <i>no</i>
<i>tobacco</i>	nominal	(1) <i>yes</i> (2) <i>no</i>
<i>contact other covid</i>	nominal	(1) <i>yes</i> (2) <i>no</i>
<i>ICU</i>	nominal	(1) <i>yes</i> (2) <i>no</i>

Identifikasi umur pasien dengan skala data rasio, sehingga dilakukan klasifikasi pada rentang usia berdasarkan kategori umur yang dikeluarkan oleh Departemen Kesehatan RI (2009) menjadi:

Tabel 2. Kategori Umur Departemen Kesehatan RI (2009)

Fitur	Rentang Umur (tahun)	Kategori
<i>age</i>	<= 5	(1) <i>toddler</i>
	6 - 11	(2) <i>children</i>
	12 - 25	(3) <i>adolescents</i>
	26 - 45	(4) <i>adult</i>
	46 - 65	(5) <i>senior</i>
	> 65	(6) <i>elderly</i>

Secara deskripsi balita (0-5 tahun), kanak-kanak (6-11 tahun), remaja (12-25 tahun), dewasa

(26-45 tahun), lansia (46-65 tahun), manula (>65 tahun).

2.4 Regresi Logistik

Regresi logistik merupakan salah satu metode analisis statistik yang digunakan untuk memodelkan hubungan satu atau lebih variabel independen terhadap variabel dependen dalam bentuk biner/kategori. Variabel independen sebagai prediktor yang digunakan merupakan variabel dalam bentuk data kontinu, kategori, atau biner [21].

Hubungan antara variabel independen terhadap variabel dependen memiliki model sebagai berikut [22]:

a. Logistic regression model

$$\log\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 X \quad (1)$$

b. Multiple regression model

$$\log\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p \quad (2)$$

2.5 Klasifikasi

Klasifikasi pada data mining dilakukan pada data yang memiliki fitur-fitur sekaligus target kelas. Klasifikasi bertujuan mengelompokkan *dataset* menjadi dua atau lebih kategori, sebagai contoh klasifikasi dilakukan pada tingkat pendapatan seseorang berdasarkan fitur umur, jenis kelamin, dan profesinya menjadi tiga kategori: tinggi, menengah, rendah [23].

Dalam melakukan klasifikasi, algoritma yang digunakan mempelajari *dataset* untuk memodelkan keterkaitan antar fitur/variabel yang ada beserta hasil klasifikasinya. Klasifikasi termasuk dalam konsep *supervised learning* karena *dataset* yang digunakan memiliki target kelas yang telah ditentukan.

2.6 Confusion Matrix

Model klasifikasi meramalkan setiap kelas setiap data memberikan hasil prediksi dalam bentuk label (positif atau negatif) [24].

$$M = \begin{pmatrix} TP & FN \\ FP & TN \end{pmatrix} \quad (3)$$

- *True Positive (TP)*, nilai prediksi positif pada data aktual positif.
- *False Positive (FP)*, nilai prediksi positif pada data aktual negatif.
- *True Negative (TN)*, nilai prediksi negatif pada data aktual negatif.
- *False Negative (FN)*, nilai prediksi negatif pada data aktual positif.

TP dan TN adalah prediksi yang benar

sedangkan TN dan FN merupakan kesalahan, sehingga dilakukan evaluasi yang didapatkan dari *confusion matrix*.

a. Akurasi

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

b. Precision

$$precision = \frac{TP}{TP + FP} \quad (5)$$

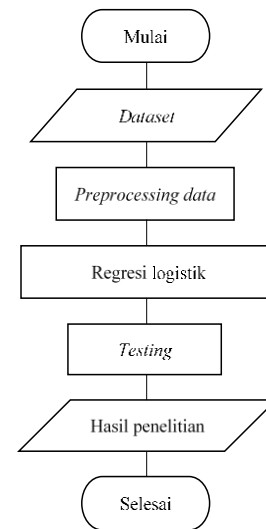
c. Recall

$$recall = \frac{TP}{TP + FN} \quad (6)$$

d. f1-score

$$f1 - score = 2 \frac{precision \cdot recall}{precision + recall} \quad (7)$$

3. Metode Penelitian



Gambar 2. Diagram Alir Penelitian

Metode penelitian merupakan proses yang dilakukan dalam melaksanakan penelitian. Langkah-langkah yang dilakukan dalam proses penelitian digambarkan dalam diagram alir pada Gambar 1. Proses penelitian dimulai dengan pengambilan data set, dilanjutkan preprocessing, kemudian proses regresi logistik dan diakhiri dengan testing. Aktivitas yang dilakukan dalam fase preprocessing diantaranya adalah: import data, data clearing, penanganan missing value, dan tranformasi data.

4. Hasil Dan Pembahasan

Berikut adalah pembahasan mengenai informasi yang didapatkan hingga hasil, serta tahapan-tahapan yang telah dilakukan sebagaimana dalam metodologi penelitian.

4.1. Data Preprocessing

Tahap ini dimulai dengan melakukan *import* untuk *library* yang dibutuhkan pada tahap

preprocessing yaitu *pandas* untuk olah data, *numpy* untuk *array* dan proses matematika, serta *matplotlib* dan *seaborn* untuk visualisasi data. Gambar 3 berikut merupakan tangkapan layar untuk proses *preprocessing*.

Data Preprocessing

```
import numpy as np
import pandas as pd

import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

Gambar 3. Data Preprocessing

Membaca data pada berkas pasien covid.csv yang kemudian disimpan dalam bentuk *data frame*. Gambar 4 berikut memuat tampilan ilustrasi data kondisi pasien COVID-19.

```
df = pd.read_csv('data/covid.csv')
df.head()
```

	id	sex	patient_type	entry_date	date_symptoms	date_died	intubed	pneumonia
0	16169f	2	1	04-05-2020	02-05-2020	9999-99-99	97	2
1	1009bf	2	1	19-03-2020	17-03-2020	9999-99-99	97	2
2	167386	1	2	06-04-2020	01-04-2020	9999-99-99	2	2
3	0b5948	2	2	17-04-2020	10-04-2020	9999-99-99	2	1
4	0d01b5	1	2	13-04-2020	13-04-2020	22-04-2020	2	2

Gambar 4. Tampilan Data Frame

Proses *data cleaning* dengan melakukan eliminasi pada fitur-fitur yang tidak diperlukan seperti kode identitas pasien, tanggal dimulai gejala, dan tanggal kematian. Selain itu memastikan bahwa tidak ditemukan data yang tidak sesuai standar seperti data laki-laki dengan kondisi mengandung (hamil). Tangkapan layar untuk data *cleaning* terdapat pada Gambar 5.

Data Cleaning

```
df = df.drop(columns=['id', 'entry_date', 'date_symptoms', 'date_died'])
df = df.replace({'sex':2, 'pregnancy':(97,98,99)}, 2)
```

Gambar 5. Data Cleaning

Menangani kondisi *missing value* pada *data frame* dengan metode eliminasi untuk baris data dengan kondisi tersebut. Nilai 97, 98, dan 99 pada *dataset* merupakan kode *missing value* seperti tidak diketahui, proses identifikasi, atau tidak tersedia. Nilai 2 dan 3 pada fitur *covid_res* merupakan kode untuk pasien negatif COVID-19 dan *missing value*. Gambaran proses penanganan *missing value* terdapat pada Gambar 6 berikut.

Menangani Missing Value

```
df = df.replace((97,98,99), np.NaN)
df = df.replace({'covid_res':(2,3)}, np.NaN)

df = df.dropna().reset_index(drop=True)
```

Gambar 6. Menangani Missing Value

Setelah penanganan *missing value*, ditemukan 2 fitur yang hanya memiliki satu nilai sehingga dilakukan tindakan eliminasi terhadap fitur sebagaimana pada Gambar 7. Fitur yang dieliminasi dari *data frame* yaitu fitur:

- *covid_res* dengan nilai 1 yang berarti seluruh data individu pasien pada *data frame* memiliki hasil positif COVID-19.
- *patient_type* dengan nilai 2 yang berarti seluruh data individu pasien pada *data frame* merupakan pasien rawat inap rumah sakit.

```
df['covid_res'].value_counts()

1.0    32639
Name: covid_res, dtype: int64
```

```
df['patient_type'].value_counts()

2    32639
Name: patient_type, dtype: int64
```

```
df = df.drop(columns=['covid_res', 'patient_type'])
```

Gambar 7. Eliminasi Fitur yang Memiliki Satu Nilai

Transformasi untuk skala data dengan tujuan seluruh fitur pada *data frame* menjadi fitur kategori. Transformasi data yang dilakukan dalam penelitian ini dilakukan pada fitur *age* dan *sex* sebagaimana Gambar 8 dan Gambar 9 berikut.

Transformasi Data

```
def age_range(age):
    if age<=5:
        return 'toddler'
    elif (age>5) and (age<=11):
        return 'children'
    elif (age>11) and (age<=25):
        return 'adolescents'
    elif (age>25) and (age<=45):
        return 'adult'
    elif (age>45) and (age<=65):
        return 'seniors'
    else:
        return 'elderly'
```

Gambar 8. Transformasi Data pada Fitur Age

```
def gender(sex):
    if sex == 1:
        return 'female'
    else:
        return 'male'
```

Gambar 9. Transformasi Data pada Fitur sex

Setelah itu dilakukan pengubahan nilai fitur pada *dataset* yang bernilai 1 dan 2 pada *data frame* untuk menyesuaikan fungsi *library* untuk tahap *modeling*, yaitu nilai 1 dan 0. Nilai 2 yang berarti "tidak", diubah menjadi nilai 0 dengan arti yang sama. Proses tersebut sebagaimana terlihat pada Gambar 10.

	sex	intubed	pneumonia	age	pregnancy	diabetes
0	female	0	0	adolescents	0	0
1	female	0	0	seniors	0	0
2	male	1	1	seniors	0	0
3	female	0	1	seniors	0	0
4	female	1	1	elderly	0	1

Gambar 10. Hasil Konversi Nilai

Dan terakhir adalah proses konversi variabel kategori pada fitur menjadi *dummy variable* atau variabel indikator sebagaimana Gambar 11.

	female	male	intubed	pneumonia	adolescents	adult	children
0	1	0	0	0	1	0	0
1	1	0	0	0	0	0	0
2	0	1	1	1	0	0	0
3	1	0	0	1	0	0	0
4	1	0	1	1	0	0	0

Gambar 11. Hasil Konversi Dummy Variable

4.2. Modeling dan Testing

Tahap melakukan *modeling* pada *dataset* dimulai dengan melakukan *import library* yang diperlukan sebagaimana Gambar 12. Selanjutnya melakukan *split* untuk *data training* dan *data testing* dengan rasio perbandingan 80% : 20% sebagaimana pada Gambar 13 dan Gambar 14.

```
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.linear_model import LogisticRegression
```

Gambar 12. Import Library sklearn

```
X = df.drop(columns=['icu'])
y = df['icu']

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.20, random_state=42)
```

Gambar 13. Modeling pada Data

```
model = LogisticRegression()
model.fit(X_train, y_train)

predictions = model.predict(X_test)

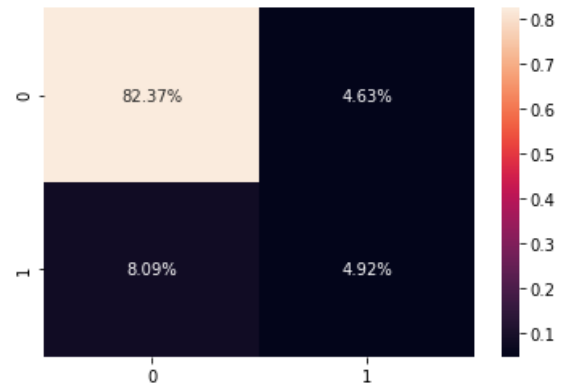
accuracy = accuracy_score(y_test, predictions)*100
print('Accuracy:', accuracy, '%', '\n')

print(classification_report(y_test, predictions))
sns.heatmap(confusion_matrix(y_test, predictions)/np.sum(
    confusion_matrix(y_test, predictions), annot=True, fmt='.2%'))
```

Gambar 14. Testing pada Data

4.3. Classification Report

Dari data hasil training dan testing yang telah dilakukan, kemudian dilakukan perhitungan *confusion matrix* sebagaimana Gambar 15. Terlihat bahwa nilai *true positive* sebesar 4,92%, *true negative* sebesar 82,37%, *false positive* sebesar 4,63%, dan *false negative* sebesar 8,09%.



Gambar 15. Confusion Matrix Hasil Klasifikasi

Berdasarkan hasil analisis pada klasifikasi admisi ruang ICU pasien COVID-19 menggunakan model regresi logistik pada confusion matrix tersebut kemudian dilakukan perhitungan akurasi menggunakan rumus (4) dan didapatkan hasil sebesar 87,29% sebagaimana Gambar 16.

Accuracy: 87.28553921568627 %

	precision	recall	f1-score	support
0	0.91	0.95	0.93	5679
1	0.52	0.38	0.44	849
accuracy			0.87	6528
macro avg	0.71	0.66	0.68	6528
weighted avg	0.86	0.87	0.86	6528

Gambar 16. Perhitungan Accuracy

4.4. Hasil Penelitian

Hasil penelitian mengenai klasifikasi pada data admisi ruang ICU untuk pasien COVID-19 menggunakan model regresi logistik menghasilkan nilai akurasi sebesar 87.29% serta rekomendasi sederhana pada Jupyter Notebook.

```
predicted = model.predict(X.loc[1:1])
if(predicted==1):
    print("Rekomendasi: ICU")
else:
    print("Rekomendasi: NON-ICU")
```

Rekomendasi: NON-ICU

Gambar 17. Rekomendasi Sederhana Admisi Ruang ICU

5. Kesimpulan

Berdasarkan hasil uji coba yang didapatkan pada implementasi model regresi logistik diperoleh perhitungan confusion matrix sebagai berikut: nilai *true positive* sebesar 4,92%, *true negative* sebesar 82,37%, *false positive* sebesar 4,63%, dan *false negative* sebesar 8,09%. Dan nilai akurasi sebesar 87,29%.

Daftar Pustaka

- [1] T. P. Velavan and C. G. Meyer, "The COVID-19 epidemic," *Trop. Med. Int. Heal.*, vol. 25, no. 3, pp. 278–280, Mar. 2020, doi: 10.1111/tmi.13383.

- [2] V. G. Puelles *et al.*, “Multiorgan and Renal Tropism of SARS-CoV-2,” *N. Engl. J. Med.*, vol. 383, no. 6, pp. 590–592, May 2020, doi: 10.1056/NEJMc2011400.
- [3] S. Adapa *et al.*, “COVID-19 Pandemic Causing Acute Kidney Injury and Impact on Patients With Chronic Kidney Disease and Renal Transplantation,” *J. Clin. Med. Res.*, vol. 12, no. 6, pp. 352–361, Jun. 2020, doi: 10.14740/jocmr4200.
- [4] C. D. C. C.-19 R. Team, “Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) - United States, February 12-March 16, 2020,” *MMWR. Morb. Mortal. Wkly. Rep.*, vol. 69, no. 12, pp. 343–346, Mar. 2020, doi: 10.15585/mmwr.mm6912e2.
- [5] M. M. Hosey and D. M. Needham, “Survivorship after COVID-19 ICU stay,” *Nat. Rev. Dis. Prim.*, vol. 6, no. 1, p. 60, 2020, doi: 10.1038/s41572-020-0201-1.
- [6] G. B. Nair and M. S. Niederman, “Updates on community acquired pneumonia management in the ICU,” *Pharmacol. Ther.*, vol. 217, p. 107663, Jan. 2021, doi: 10.1016/j.pharmthera.2020.107663.
- [7] L. Roncon, M. Zuin, G. Rigatelli, and G. Zuliani, “Diabetic patients with COVID-19 infection are at higher risk of ICU admission and poor short-term outcome,” *J. Clin. Virol.*, vol. 127, p. 104354, 2020, doi: 10.1016/j.jcv.2020.104354.
- [8] J. M. Leung, M. Niikura, C. W. T. Yang, and D. D. Sin, “COVID-19 and COPD,” *Eur. Respir. J.*, vol. 56, no. 2, p. 2002108, Aug. 2020, doi: 10.1183/13993003.02108-2020.
- [9] D. M. G. Halpin *et al.*, “Global Initiative for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease. The 2020 GOLD Science Committee Report on COVID-19 and Chronic Obstructive Pulmonary Disease,” *Am. J. Respir. Crit. Care Med.*, vol. 203, no. 1, pp. 24–36, Nov. 2020, doi: 10.1164/rccm.202009-3533SO.
- [10] E. M. Abrams, G. W. ‘t Jong, and C. L. Yang, “Asthma and COVID-19,” *Can. Med. Assoc. J.*, vol. 192, no. 20, pp. E551–E551, May 2020, doi: 10.1503/cmaj.200617.
- [11] Y. Gao, Y. Chen, M. Liu, S. Shi, and J. Tian, “Impacts of immunosuppression and immunodeficiency on COVID-19: A systematic review and meta-analysis,” *J. Infect.*, vol. 81, no. 2, pp. e93–e95, Aug. 2020, doi: 10.1016/j.jinf.2020.05.017.
- [12] R. Pranata, M. A. Lim, I. Huang, S. B. Raharjo, and A. A. Lukito, “Hypertension is associated with increased mortality and severity of disease in COVID-19 pneumonia: A systematic review, meta-analysis and meta-regression,” *J. Renin-Angiotensin-Aldosterone Syst.*, vol. 21, no. 2, p. 147032032092689, Apr. 2020, doi: 10.1177/1470320320926899.
- [13] K. J. Clerkin *et al.*, “COVID-19 and Cardiovascular Disease,” *Circulation*, vol. 141, no. 20, pp. 1648–1655, May 2020, doi: 10.1161/CIRCULATIONAHA.120.046941.
- [14] F. Gao *et al.*, “Obesity Is a Risk Factor for Greater COVID-19 Severity,” *Diabetes Care*, vol. 43, no. 7, p. e72 LP-e74, Jul. 2020, doi: 10.2337/dc20-0682.
- [15] R. Patanavanich and S. A. Glantz, “Smoking Is Associated With COVID-19 Progression: A Meta-analysis,” *Nicotine Tob. Res.*, vol. 22, no. 9, pp. 1653–1656, Aug. 2020, doi: 10.1093/ntr/ntaa082.
- [16] S. Ellington *et al.*, “Characteristics of Women of Reproductive Age with Laboratory-Confirmed SARS-CoV-2 Infection by Pregnancy Status - United States, January 22-June 7, 2020,” *MMWR. Morb. Mortal. Wkly. Rep.*, vol. 69, no. 25, pp. 769–775, Jun. 2020, doi: 10.15585/mmwr.mm6925a1.
- [17] S. M. Azizah and N. E. Chandra, “MODEL REGRESI LOGISTIK PADA FAKTOR-FAKTOR YANG MEMPENGARUHI IMUNISASI LENGKAP BALITA,” *J. Ilm. Teknosains*, vol. 3, no. 2, Nov. 2017, doi: 10.26877/jitek.v3i2.1882.
- [18] S. H. Adil, M. Ebrahim, K. Raza, S. S. Azhar Ali, and M. Ahmed Hashmani, “Liver Patient Classification using Logistic Regression,” in *2018 4th International Conference on Computer and Information Sciences (ICCOINS)*, Aug. 2018, pp. 1–5, doi: 10.1109/ICCOINS.2018.8510581.
- [19] M. Saw, T. Saxena, S. Kaithwas, R. Yadav, and N. Lal, “Estimation of Prediction for Getting Heart Disease Using Logistic Regression Model of Machine Learning,” in *2020 International Conference on Computer Communication and Informatics (ICCCI)*, 2020, pp. 1–6, doi: 10.1109/ICCCI48352.2020.9104210.
- [20] R. J. Roiger, *Data Mining*, 2nd Editio. Chapman and Hall/CRC, 2017.
- [21] J. M. Hilbe, *Logistic Regression Models*. Chapman and Hall/CRC, 2009.
- [22] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, vol. 103. New York, NY: Springer New York, 2013.
- [23] D. T. Larose and C. D. Larose, *Discovering knowledge in data: an introduction to data mining*, vol. 4. John Wiley & Sons, 2014.
- [24] D. Chicco and G. Jurman, “The advantages of

the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation,” *BMC Genomics*, vol. 21, no. 1, p. 6, 2020, doi: 10.1186/s12864-019-6413-7.