

Permodelan Regresi Logistik untuk Diagnosis Dini Infeksi Covid-19 Varian Delta di Indonesia

Nila Ayu Nur Roosyidah¹, Putu Krishnanda Supriyatna^{2*}

^{1,2}Jurusan Statistika, Fakultas Sains dan Analitika Data, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia

*Corresponding author. Email: nila.roosyidah12@gmail.com

ABSTRAK

Untuk mengendalikan penyebaran virus di Indonesia, pemerintah terus mengupayakan pelaksanaan 3T (*Testing, Training, dan Tracing*) secara menyeluruh. Pemberlakuan *testing* secara masif seringkali terkendala oleh beberapa hal, di antaranya biaya dan keterjangkauan akses. Penelitian ini bertujuan untuk membuat sebuah model untuk diagnosis dini kasus infeksi melalui beberapa gejala penciri dan pengalaman kontak erat dengan pasien positif. Data yang digunakan merupakan data primer yang diperoleh melalui survei *online* dengan teknik pengambilan sampel *non probability*. Kemudian data yang diperoleh dianalisis menggunakan model regresi logistik biner dan diperoleh hasil bahwa gejala anosmia, gejala meriang, dan pengalaman kontak erat signifikan dalam memengaruhi infeksi Covid-19. Dari nilai *odds ratio*, diketahui anosmia merupakan variabel yang paling berpengaruh. Seseorang yang mengalami anosmia memiliki kecenderungan untuk terinfeksi sebesar 31 kali dibandingkan yang tidak mengalami. Validasi kekuatan model dalam melakukan klasifikasi dilakukan dengan melakukan prediksi pada model yang dihasilkan sudah baik, karena pengukuran masing-masing kriteria kekuatan model yang terdiri dari akurasi, sensitivitas, dan spesifisitas model baik pada data *testing* masing-masing menghasilkan nilai 0,80 (telah mendekati 1). Luasan area di bawah kurva *Receiver Operating Characteristic* (ROC) untuk data *testing* adalah 0,85 yang berarti bahwa model sudah memiliki kriteria baik (*good*) dalam melakukan klasifikasi.

Kata Kunci:

Covid-19; Diagnosis Dini; Regresi Logistik, *Receiver Operating Characteristic*

ABSTRACT

Controlling the spread of the Covid-19 virus in Indonesia, the government continues to strive for a comprehensive 3T (*Testing, Training, and Tracing*) implementation. Massive testing is often constrained by several things, including cost and affordability of access. This study aims to create a model for early diagnosis of Covid-19 infection cases through several characteristic symptoms and experiences of close contact with positive patients. By using a binary logistic regression model, it was found that symptoms of anosmia, feverish symptoms, and close contact experience were significant in influencing Covid-19 infection. From the odds ratio value, it is known that anosmia is the most influential variable. Someone who has anosmia tends to be infected by 31 times compared to those who do not. Validation of the strength of the model in classifying is done by making predictions on the resulting model is good, because the measurement of each criterion of the strength of the model consists of accuracy, sensitivity, and specificity of the model both on the data testing each produces a value of 0.8 (close to 1). The area under the Receiver Operating Characteristic (ROC) curve for testing data is 0.8462, which means that the model already has good criteria for classifying.

Keywords:*Covid-19; Early Diagnose; Logistic Regression; Receiver Operating Characteristic***Format Sitasi:**

N. A. N Roosyidah and P. K. Supriyatna, "Permodelan Regresi Logistik untuk Diagnosis Dini Infeksi Covid-19 Varian Delta di Indonesia", *Jambura J. Math.*, vol. 4, No. 2, pp. 232–246, 2022, doi: <https://doi.org/10.34312/jjom.v4i2.12653>

1. Pendahuluan

Hingga Bulan Desember 2021, pandemi Covid-19 telah berlangsung kurang lebih 2 tahun dengan jumlah kasus terkonfirmasi positif di dunia mencapai lebih dari 264 juta jiwa atau sebanyak 264.456.031 dengan jumlah kasus meninggal sebesar 25.518.505, dan kasus sembuh sebanyak 238.937.526. Tidak sedikit dampak yang dirasakan oleh negara-negara terdampak, salah satunya Indonesia. Kondisi pandemic meyebabkan perlambatan ekonomi akibat imbas dari rendahnya sentimen investor terhadap pasar [1].

Berbagai upaya telah dilakukan pemerintah dan segenap jajarannya untuk mengatasi dampak pandemi Covid-19, salah satunya melalui upaya memutus mata rantai persebaran dengan mekanisme 3T (*Testing, Tracing, dan Treatment*) [2]. *Testing* merupakan tindakan pemeriksaan kondisi seseorang untuk mengetahui apakah telah terinfeksi virus Covid-19 melalui pengujian laboratorium baik dengan cara pengujian usap (swab) antigen maupun PCR (*polymerase chain reaction*). *Tracing* merupakan upaya menelusuri Riwayat kontak dari pasien positif Covid-19. Terakhir, *treatment* merupakan Tindakan perawatan dari pasien yang telah terkonfirmasi positif Covid-19, baik dengan isolasi mandiri di rumah maupun perawatan di rumah sakit.

Pelaksanaan *testing* secara masif tentu sangat efektif karena dapat menjangkit sebanyak-banyaknya kasus positif untuk dilakukan tindakan *tracing* dan *treatment* sehingga penyebaran wabah dapat dikendalikan. Akan tetapi, terdapat beberapa hambatan dalam pelaksanaan test Covid-19. Pertama, biaya yang cukup besar untuk melakukan test, baik swab antigen maupun swab RT-PCR. Meskipun pada akhirnya pemerintah menurunkan tarif pemeriksaan RT-PCR menjadi 275 ribu untuk wilayah Jawa Bali dan 300 ribu untuk daerah lainnya, pada kenyataannya besaran tersebut masih cukup menjadi beban bagi sebagian masyarakat [3]. Andaikan beban tes dibebankan kepada pemerintah pun tidak akan menjadi solusi karena penambahan pengeluaran pemerintah berdampak pada penambahan beban APBN yang hingga saat ini sebagian masih didanai oleh utang pemerintah.

Hambatan kedua adalah tidak meratanya fasilitas kesehatan yang menyediakan pemeriksaan Covid-19 di seluruh Indonesia. Hal ini tentu menyebabkan delay waktu yang cukup lama untuk mengetahui hasil laboratorium. Belum lagi pengiriman yang memerlukan waktu yang cukup lama memungkinkan specimen uji terkontaminasi sehingga meningkatkan risiko terjadinya *false negativity result* [4].

Ketiga, hambatan pelaksanaan *testing* Covid-19 secara masif berasal dari sisi pasien *suspect* sendiri. Hal ini pada umumnya dipicu oleh ketakutan atas stigma negatif dari lingkungan sekitar apabila hasil dari pemeriksaan menyatakan positif Covid-19. Agustang, et.al [5] menemukan bahwa dampak buruk lainnya dari pandemic Covid-19 adalah munculnya stigma negatif terhadap pasien Covid-19 yang bahkan terus melekat

meskipun pasien telah meninggal dunia.

Vaksinasi sebagai salah satu cara yang diambil pemerintah untuk menangani pandemi Covid-19 diharapkan dapat memperkuat tubuh sehingga lebih mudah disembuhkan jika terjadi infeksi. Namun, Dedi Junaidi menemukan bahwa vaksinasi justru signifikan dalam meningkatkan jumlah kasus Covid-19 beserta kematiannya. Namun di sisi lain, angka kesembuhan juga meningkat drastis dan jauh lebih signifikan dibandingkan peningkatan jumlah kasus dan kematian [6]. Karena pada penelitian ini tidak meneliti angka kesembuhan, maka akan meneliti apakah program vaksinasi memiliki efek samping dalam meningkatkan jumlah kasus Covid-19. Selain itu, adanya penyakit bawaan pada responden diduga mampu membuat responden lebih terserang virus. CDC (*Centers for Disease Control and Prevention*) pada tahun 2020 melaporkan bahwa 94 persen pasien Covid-19 meninggal karena memiliki penyakit bawaan [1]. Adanya penyakit bawaan tersebut membuat tubuh lemah sehingga mudah terinfeksi oleh virus.

Gaya hidup yang kurang sehat seperti kurang tidur, kebiasaan merokok, dan jam kerja yang cukup panjang bisa juga menjadi penyebab seseorang mudah terjangkit penyakit. Covid-19 merupakan penyakit yang menginfeksi sistem pernafasan. Para perokok dianggap lebih rentan terjangkit virus ini. Laporan dari *World Health Organization* (WHO) menyebutkan bahwa perokok lebih beresiko tinggi untuk masuk ruang ICU dan memiliki risiko kematian yang tinggi ketika terjangkit virus Covid-19 dibandingkan yang bukan perokok [7]. Pandemi Covid-19 merubah pola tidur kebanyakan orang. Semenjak pandemic berlangsung, prevalensi insomnia pun mengalami peningkatan. Kebiasaan begadang dikhawatirkan akan menurunkan sistem imunitas secara jangka panjang. Kurang istirahat, jam kerja panjang, serta tekanan kerja yang tinggi pada pelayanan kesehatan menyebabkan banyaknya jatuhnya korban di bidang tersebut akibat pandemi Covid-19 [8].

Indikator utama yang bisa digunakan untuk mengidentifikasi suatu penyakit adalah gejala yang dialami pasien. Gejala-gejala tersebut antara lain anosmia, demam, batuk terus menerus, meriang, kehilangan nafsu makan, sakit pada otot atau persendian [9]. Selain itu, ditemukan juga bahwa anosmia merupakan salah satu manifestasi klinis paling umum pada penderita wabah SARS-CoV-2 yang diawali dengan tidak berfungsinya kemosensorik pada indra yang melibatkan rasa dan penciuman [10]. Selain anosmia, demam dan meriang juga sering menjadi gejala Covid-19 [11]. Gejala lainnya yang ditemukan adalah kelelahan luar biasa dan sesak nafas [2].

Mobilitas juga dianggap sebagai variabel penting yang memengaruhi kejadian infeksi Covid-19 karena virus ini bisa menyebar ke seluruh dunia tak terlepas dari mobilitas penduduk. Penyebaran virus Covid-19 bisa terjadi antarmanusia secara langsung maupun tidak langsung. Penyebaran virus secara langsung diantaranya terjadi melalui kontak erat dengan orang yang terinfeksi melalui sekresi air liur, sekresi pernafasan, maupun dari percikan sekresi (droplet) yang menyebar melalui udara. Sementara itu, penyebaran tidak langsung dapat terjadi melalui permukaan benda yang terkontaminasi oleh virus [12]. Dengan banyaknya metode penularan tersebut, maka peluang untuk tertular sangat besar. Dalam penelitian yang lain juga ditemukan bahwa mobilitas manusia terbukti berpengaruh terhadap tingkat penyebaran sehingga sering diangkat sebagai faktor penting dalam upaya pengendalian penyebaran penyakit menular [13].

Regresi logistik merupakan metode analisis untuk mengetahui hubungan antara

variabel tak bebas (*response variabel*) yang memiliki dua kategori atau lebih dengan satu atau lebih variabel bebas (*independent variabels*), baik yang bersifat kategorik maupun numerik. Melalui metode tersebut akan diketahui seberapa mampu variabel bebas dalam menjelaskan keragaman pada variabel tak bebas di dalam model [6]. Penggunaan regresi logistik untuk deteksi dini penyakit pernah dilakukan sebelumnya untuk mendeteksi penyakit kanker paru-paru melalui sampel udara yang dihembuskan melalui pernafasannya [14]. Selain itu, diagnosis penyakit Alzheimer pernah juga dilakukan menggunakan model *sparse logistic regression* [15]. Selanjutnya, deteksi dini kejadian infeksi virus Covid-19 pernah dilakukan oleh Song, et.al [16] di Rumah Sakit Zhejicang, Huangzhou, untuk membuat sebuah *tool* untuk mengidentifikasi pasien yang lebih utama dicurigai sebagai penderita (*highly suspected patients*).

Penelitian ini memiliki tujuan untuk menghasilkan sebuah model statistik yang dapat digunakan sebagai alternatif untuk melakukan deteksi dini apakah seseorang telah terjangkit virus Covid-19 berdasarkan profil, gaya hidup, gejala yang dialami, dan mobilitasnya. Model ini juga diharapkan dapat membantu pemerintah dalam menyeleksi siapa yang lebih diutamakan untuk dilakukan *testing*. Deteksi dini ini diharapkan bisa menjadi alternatif ketika menghadapi kendala untuk melakukan tes Covid-19 secara cepat. Bercermin dari parahnya dampak dari gelombang varian delta sebelumnya yang terjadi di sekitar Bulan Juni hingga Agustus, maka penelitian ini memilih bulan tersebut sebagai waktu referensi.

2. Metode

2.1. Tahapan Penelitian

Data yang digunakan dalam penelitian ini merupakan data primer yang didapatkan melalui survei *online* menggunakan kuesioner yang dibuat menggunakan aplikasi *google* formulir dengan metode pemilihan sampel *non probability*. Cakupan responden adalah individu yang pernah melakukan tes Covid-19, baik *swab* RT-PCR maupun antigen selama bulan Juni hingga Agustus 2021. Alasan pemilihan waktu tersebut adalah karena pada saat itu Indonesia sedang dilanda gelombang varian delta dengan cukup parah.

Responden yang diperoleh sejumlah 107 yang kemudian dikelompokkan menjadi data *training* sebanyak 87 dan data *testing* sebanyak 20. Data *training* diperlukan untuk membangun model sedangkan data *testing* diperlukan untuk mengetahui kekuatan dari model yang dihasilkan.

Penelitian ini menggunakan 2 metode analisis, yaitu analisis deskriptif menggunakan grafik dan tabulasi silang serta analisis inferensia menggunakan analisis regresi logistik. Analisis deskriptif bertujuan menggambarkan karakteristik responden penelitian sedangkan analisis inferensia bertujuan untuk mengetahui bagaimana pengaruh dari variabel-variabel prediktor dalam menentukan seseorang positif terjangkit virus Covid-19 atau tidak. Sementara itu, analisis inferensia digunakan untuk membangun model deteksi dini dan mengetahui faktor apa saja yang memengaruhi secara nyata (signifikan) pada terjadinya infeksi virus Covid-19. Analisis inferensia dilakukan menggunakan metode analisis regresi logistic.

Sebelum melakukan analisis regresi logistik, terlebih dahulu dilakukan pemeriksaan hubungan antara variabel prediktor dengan variabel respon menggunakan uji independensi *chi-square*. Hanya variabel yang berkorelasi signifikan yang digunakan

untuk *fitting model* regresi logistik Pemilihan model terbaik dilakukan dengan hanya mempertahankan variabel yang berpengaruh signifikan terhadap model.

2.2. Regresi Logistik

Regresi Logistik merupakan metode analisis statistik yang digunakan untuk mempelajari hubungan anatara variabel prediktor dengan variabel respon, dimana variabel respon memiliki dua kategori atau lebih. Dalam penelitian ini, regresi logistik berguna untuk memodelkan variabel apa saja yang berpengaruh signifikan pada kejadian positif terinfeksi Covid-19.

Secara umum berikut persamaan matematis model regresi logistik:

$$P(y | x) = \pi(x) \quad (1)$$

dengan

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)} \quad (2)$$

dimana y bernilai 1 untuk kejadian yang dinyatakan sukses dan 0 untuk kejadian yang dinyatakan gagal.

Oleh karena model logit merupakan bentuk nonlinear, maka diperlukan transformasi logit untuk mendapatkan model yang linear sehingga diperoleh model sebagai berikut:

$$g(x) = \ln \frac{\pi(x)}{1 - \pi(x)} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (3)$$

dengan $\frac{\pi(x)}{1 - \pi(x)}$ merupakan risiko kejadian ketika $y = 1$ untuk x bernilai tertentu, sehingga diperoleh persamaan:

$$y = \pi(x) + e \quad (4)$$

Dengan e berdistribusi binomial dengan mean 0 dan varians $(\pi(x)(1 - \pi(x)))$.

Tidak seperti persamaan regresi OLS, model regresi logistik tidak memerlukan persyaratan ketat pemenuhan asumsi klasik, di antaranya hubungan variabel bebas dengan variabel tak bebas tidak harus linier dan error tidak harus berdistribusi normal serta memiliki varians yang konstan (homoskedastisitas).

Untuk menilai keberartian model, pada model regresi logistik digunakan uji kebaikan model (*the goodness of fit test*) menggunakan pengujian *Hosmer and Lemeshow* dan penggunaan tabel klasifikasi (*confusion matrix*). Dari *confusion matrix* kemudian diperoleh 3 ukuran kekuatan model, yaitu akurasi, sensitifitas, dan spesifisitas yang menunjukkan tingkat kebenaran dari klasifikasi yang dilakukan.

Kemudian, pengaruh dari keseluruhan variabel bebas terhadap variabel tak bebas digunakan pengujian simultan menggunakan *likelihood ratio test*. Berikut merupakan hipotesis pengujian dan statistik uji yang digunakan.

$$H_0 = \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 = \text{minimal ada satu nilai } \beta_k \neq 0$$

dengan: $k = 1, 2, \dots, p$

Rumus dari statistik uji G^2 adalah:

$$G^2 = -2 \ln \left[\frac{l_0}{l_k} \right] \quad (5)$$

dimana

$l_0 = \text{likelihood}$ tanpa variabel bebas

$l_k = \text{likelihood}$ dengan variabel bebas

Oleh karena statistik uji G^2 berdistribusi χ^2 dengan derajat bebas p , maka wilayah kritis pengujian adalah jika nilai statistik uji $G > \chi^2_{p(\alpha)}$.

Untuk mengetahui pengaruh dari masing-masing predictor, dilakukan uji secara parsial menggunakan statistik uji Wald dengan rumus sebagai berikut.

$$W = \frac{\hat{\beta}_k}{se(\hat{\beta}_k)} \quad (6)$$

W mengikuti distribusi normal standar, sehingga wilayah kritis pengujian adalah saat $W < Z_{\frac{1}{2}\alpha}$ atau $W > Z_{\frac{1}{2}\alpha}$.

2.3. Evaluasi Model

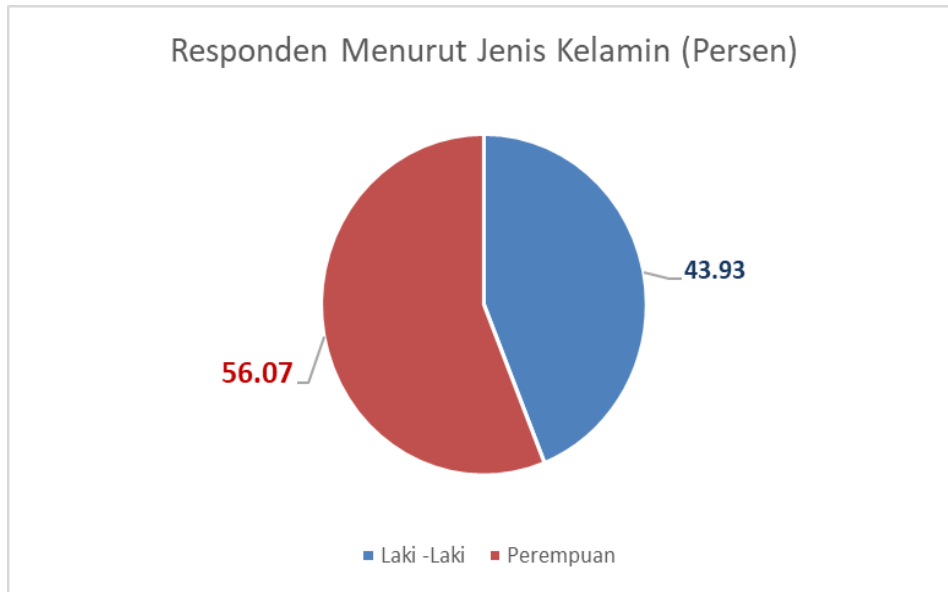
Untuk mengetahui apakah kekuatan model regresi logistik untuk memprediksi data di luar model, maka dilakukan validasi menggunakan informasi data *testing*. Kemudian dilakukan perbandingan antara hasil prediksi model dan pengelompokan aktualnya, sehingga bisa dibuat sebuah tabel klasifikasi (*confusion matrix*). Dari *confusion matrix* tersebut dapat dihitung 3 kekuatan model, yang terdiri dari akurasi, sensitivitas, dan spesifisitas dari model yang dihasilkan. Apabila ketiga ukuran tersebut mendekati 1, maka dapat dinyatakan klasifikasi telah dilakukan dengan baik.

3. Hasil dan Pembahasan

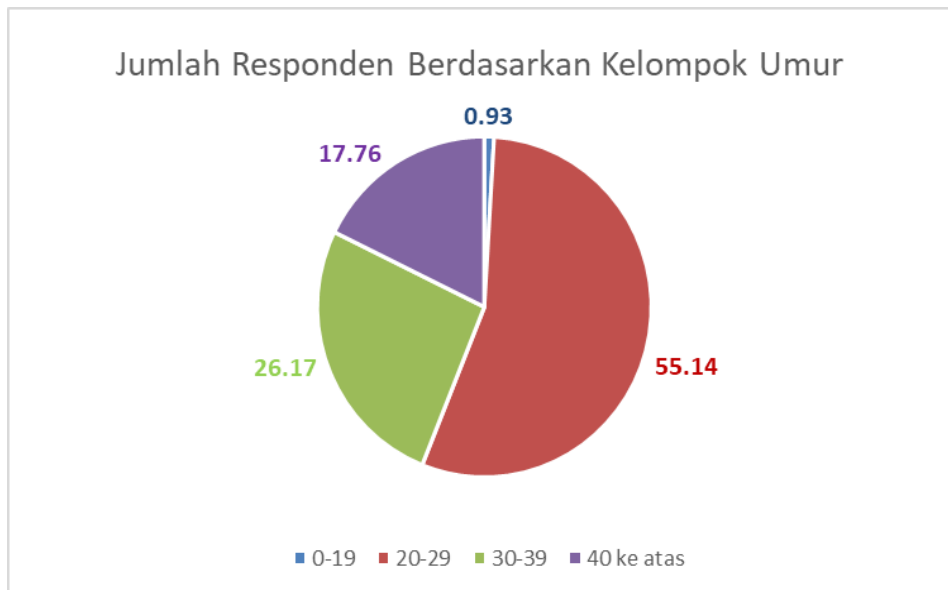
3.1. Deskripsi Responden

Berikut gambaran umum dari responden pada penelitian ini. Dari 107 responden yang digunakan pada penelitian ini, 56,07 persennya atau sebanyak 60 orang berjenis kelamin perempuan dan sisanya laki-laki seperti yang dijelaskan pada gambar 1. Kemudian, berdasarkan kelompok umur, lebih dari 55 persen responden berada pada kelompok umur 20-29 tahun, diikuti 26,17 persen berada pada kelompok umur 30-39 tahun, dan sisanya, 17,76 persen merupakan mereka yang berusia 40 tahun ke atas seperti yang tersaji pada Gambar 2.

Dari 107 responden, 38 responden merupakan pasien yang terkonfirmasi positif Covid-19. Kemudian, dilihat dari status vaksinnya, sebanyak 72,65 persen responden telah menerima vaksin 2 dosis dan jauh lebih sedikit yang pernah terinfeksi. Sebaliknya, responden yang belum pernah menerima vaksinasi Covid-19 atau sekitar 18,80 persen, lebih banyak yang terkonfirmasi positif.



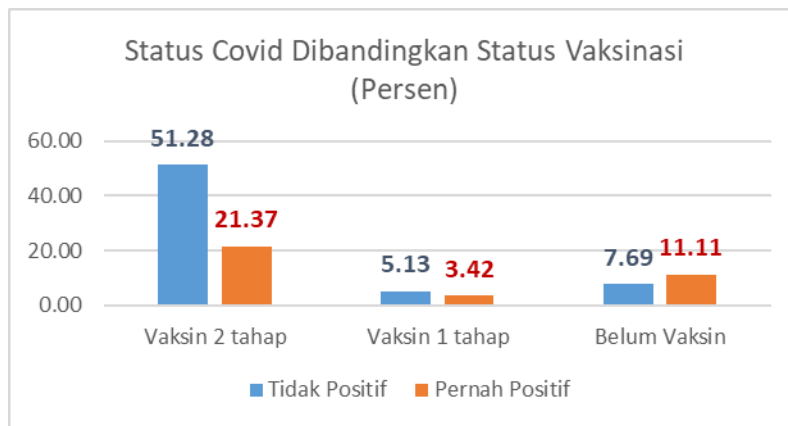
Gambar 1. Persentase responden menurut jenis kelamin



Gambar 2. Persentase responden menurut kelompok umur

Tabel 1 merangkum gejala-gejala yang dialami oleh seluruh responden. Secara umum bisa dilihat bahwa gejala tersebut lebih banyak dialami oleh responden yang pernah terkonfirmasi positif. Sebagai contoh dari 27 responden yang mengalami anosmia, 26 di antaranya merupakan responden yang pernah terinfeksi. Begitu juga dengan sesak nafas, dari 14 responden yang mengalaminya, seluruhnya pernah terkonfirmasi positif Covid-19.

Tabel 2 di bawah menunjukkan jumlah responden berdasarkan pernah atau tidaknya melakukan kontak dengan pasien Covid-19. Sebanyak 76,32 persen responden yang positif sebelumnya telah melakukan kontak erat. Sementara itu, sebanyak 62,32 persen responden yang memiliki hasil tes negatif ternyata tidak pernah melakukan kontak erat dengan pasien. Hal ini mengindikasikan bahwa kejadian terinfeksi virus Covid-19



Gambar 3. Persentase responden berdasarkan status tes Covid-19 dan status vaksinasi

Tabel 1. Jumlah responden berdasarkan status Covid-19 dan gejala yang dialami

Gejala	Status Covid-19		
	Negatif	Positif	Total
Anosmia	1	26	27
Meriang	6	24	30
Nyeri otot	9	21	30
Kelelahan	7	19	26
Hilang selera makan	3	15	28
Sesak nafas	0	14	14
Batuk	5	15	20
Demam	13	31	44

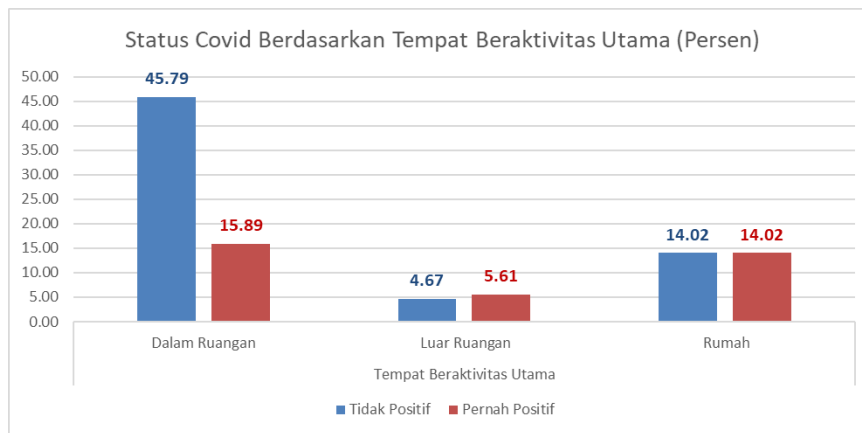
berkaitan dengan pernah atau tidaknya terjadi kontak erat dengan pasien positif.

Tabel 2. Persentase responden berdasarkan status Covid-19 dan kontak erat dengan pasien positif Covid-19 (Persen)

Status Covid-19	Kontak Erat		Total
	Pernah	Tidak Pernah	
Positif	76,32	23,68	100
Negatif	37,68	62,32	100

Gambar 4 menyajikan persentase responden berdasarkan lokasi aktivitas utamanya. Responden yang lebih banyak beraktivitas di dalam ruangan ternyata lebih sedikit yang terkonfirmasi positif Covid-19. Sementara itu, responden yang beraktivitas di luar ruangan lebih banyak yang memiliki hasil tes negatif. Hal tersebut mengindikasikan bahwa kecenderungan seseorang untuk terkena infeksi virus Covid-19 ketika bekerja di dalam ruangan tertutup semakin besar.

Gambar 4 menyajikan persentase responden berdasarkan lokasi aktivitas utamanya. Responden yang bekerja di dalam ruangan memiliki persentase tertinggi tertular Covid-19, setelahnya adalah responden yang mayoritas waktunya beraktivitas di dalam rumah. Hal tersebut mengindikasikan bahwa penularan virus Covid-19 banyak terjadi di dalam ruangan tertutup, baik di perkantoran, rumah sakit, puskesmas, dan tempat



Gambar 4. Persentase responden berdasarkan status Covid-19 dan tempat beraktivitas utamanya

lainnya. Selain itu, terdapat kemungkinan bahwa seseorang yang tertular karena bekerja di dalam ruangan tertutup tersebut menularkan kepada anggota keluarganya yang berada di rumah sehingga kasus positif Covid-19 orang yang banyak menghabiskan waktu dengan beraktivitas di dalam rumah juga tinggi.

3.2. Pengujian Independensi

Sebelum melakukan permodelan menggunakan regresi logistik biner dilakukan pemeriksaan korelasi antara variabel prediktor yang digunakan dengan variabel respon menggunakan uji independensi chi-square. Masing-masing prediktor harus memiliki hubungan yang kuat dengan variabel respons sehingga dapat dilakukan permodelan atau dengan kata lain hasil uji independensinya menghasilkan *p-value* yang lebih rendah dari tingkat signifikansi dalam penelitian. Pengujian ini menggunakan signifikansi (α) 5%. Tabel 3 merupakan hasil pengujian independensi untuk masing-masing variabel.

Tabel 3. Uji Independensi

Variabel	χ^2_{hitung}	Df	P-value	Keterangan
Vaksin (D_1)	7,28	2	0,03	Signifikan
Memiliki comorbid (D_2)	0,16	1	0,70	Tidak signifikan
Perokok (D_3)	0,26	1	0,61	Tidak signifikan
Frekuensi Tidur (D_4)	$1,58 \times 10^{-29}$	1	1	Tidak signifikan
Kepadatan Aktivitas (D_5)	0,10	1	0,74	Tidak signifikan
Anosmia (D_6)	54,76	1	$1,36 \times 10^{-3}$	Signifikan
Demam (D_7)	37,29	1	$1,02 \times 10^{-3}$	Signifikan
Batuk (D_8)	14,69	1	$1,27 \times 10^{-4}$	Signifikan
Sesak Nafas (D_9)	26,10	1	$3,25 \times 10^{-7}$	Signifikan
Hilang Selera Makan (D_{10})	19,17	1	$1,20 \times 10^{-5}$	Signifikan
Kelelahan (D_{11})	19,05	1	$1,27 \times 10^{-5}$	Signifikan
Nyeri Otot (D_{12})	19,61	1	$9,51 \times 10^{-6}$	Signifikan
Meriang (D_{13})	33,38	1	$7,60 \times 10^{-9}$	Signifikan
Kontak (D_{14})	13,14	1	$2,9 \times 10^{-4}$	Signifikan
Workplace (D_{15})	2,99	2	0,08	Tidak Signifikan

Berdasarkan Tabel 3, diketahui bahwa variabel memiliki comorbid, perokok, frekuensi

tidur, kepadatan aktivitas, dan tempat kerja (*workplace*) tidak signifikan berkorelasi dengan variabel status positif dan negatif tes Covid-19. Untuk itu, pada tahapan pembentukan model selanjutnya, variabel tersebut tidak diikut sertakan.

3.3. Pemodelan Klasifikasi Status Positif dan Negatif Test Covid-19 dengan Regresi Logistik Biner

Setelah melakukan pemeriksaan hubungan antara variabel prediktor dengan variabel respons, tahap selanjutnya adalah membangun model regresi logistik biner menggunakan prediktor yang memiliki korelasi signifikan dengan variabel respons. Hasil estimasi parameter tersaji dalam Tabel 4.

Tabel 4. Estimasi Parameter model regresi logistik biner

Parameter	$\widehat{(\beta)}$	Z-value	p-value
Intersep	-2,26	-1,47	0,14
Status Vaksinasi 1 Dosis $\widehat{(\beta)}_{11}$	-2,56	-1,18	0,24
Status Vaksinasi 2 Dosis $\widehat{(\beta)}_{12}$	-1,85	-1,46	0,15
Anosmia $\widehat{(\beta)}_{21}$	35,47	0,01	0,99
Demam $\widehat{(\beta)}_{31}$	-1,33	-0,84	0,40
Batuk $\widehat{(\beta)}_{41}$	-16,78	-4×10^{-3}	0,99
Sesak nafas $\widehat{(\beta)}_{51}$	56,22	7×10^{-3}	0,99
Hilang selera makan $\widehat{(\beta)}_{61}$	0,87	0,54	0,59
Kelelahan $\widehat{(\beta)}_{71}$	-17,66	-4×10^{-3}	0,99
Nyeri otot $\widehat{(\beta)}_{81}$	1,18	0,82	0,41
Meriang $\widehat{(\beta)}_{91}$	3,85	2,74	$6,1 \times 10^{-3}$ **
Kontak $\widehat{(\beta)}_{101}$	2,09	1,59	0,11

Keterangan: **) signifikan pada alpha 5%

Model yang dihasilkan pada *running* pertama memiliki nilai statistic uji $G^2 = 79,38$ dengan *p-value* mendekati 0 atau sangat kecil sekali. Pada signifikansi 5% maka dapat dinyatakan bahwa secara simultan seluruh prediktor secara bersama-sama memengaruhi penentuan klasifikasi status tes Covid-19. Kemudian, nilai statistik uji pada Hosmer-Lemeshow adalah 3,09 dengan *p-value* 0,93 sehingga diperoleh kesimpulan bahwa model tersebut *fit* atau dapat digunakan untuk melakukan klasifikasi. Dari model juga dapat diketahui nilai *Akaike Information Criteria* (AIC) sebesar 57,94. Akan tetapi, pengujian secara parsial menggunakan Wald Test menghasilkan hanya gejala meriang yang berpengaruh signifikan terhadap penentuan positif atau negatifnya hasil tes Covid-19 seperti yang ditampilkan pada Tabel 4.

Oleh karena tujuan dari pembentukan model adalah pembuatan klasifikasi, maka hanya variabel-variabel yang berpengaruh signifikan yang dipertahankan di dalam model. Seleksi variabel dilakukan melalui cara memasukkan satu per satu variabel untuk melihat variabel mana saja yang signifikan. Variabel lain yang tidak signifikan akan dikeluarkan dari model (metode *Forward Wald*). Adapun hasil estimasi parameternya disajikan pada Tabel 5.

Pengujian secara serentak menggunakan *likelihood ratio test* menghasilkan statistik uji 64,35 dengan *p-value* < 5%. Artinya, seluruh prediktor secara bersama-sama memengaruhi penentuan status test Covid-19. Kemudian, pengujian secara parsial pada masing-masing variabel anosmia, meriang, dan kontak menghasilkan *p-value* kurang dari 5%. Sehingga dapat diperoleh kesimpulan bahwa masing-masing variabel tersebut

Tabel 5. Estimasi Parameter model regresi logistic biner model terbaik

Parameter	$\widehat{(\beta)}$	$Exp(\widehat{\beta})$	Z-value	p-value **
Intersep	-3,53	0,03	-4,11	$3,94 \times 10^{-5}$
Anosmia $(\widehat{\beta})_{21}$	3,44	31,29	3,04	$2,35 \times 10^{-3}$
Meriang $(\widehat{\beta})_{91}$	2,56	12,97	2,82	$4,82 \times 10^{-3}$
Kontak $(\widehat{\beta})_{101}$	2,15	8,62	2,41	0,02

Keterangan: **) signifikan pada alpha 5%

pada tingkat kepercayaan 95% memengaruhi penentuan klasifikasi seseorang untuk dinyatakan sebagai pasien positif Covid-19 atau tidak. Pengujian kebaikan model menggunakan Hosmer Lemeshow Test menghasilkan p-value lebih dari $\alpha = 5\%$ sehingga model ini dikatakan baik (fit) dalam melakukan klasifikasi.

AIC merupakan kriteria yang dapat menunjukkan seberapa sesuai hasil klasifikasi yang dilakukan oleh model dengan kondisi yang sebenarnya. Semakin kecil nilainya maka peluang ketepatan klasifikasi oleh model semakin besar. Meskipun model yang memiliki AIC lebih kecil adalah model yang lebih sederhana, model tersebut hendaknya yang lebih dipilih karena memiliki kecenderungan untuk lebih mirip dengan nilai yang sebenarnya [17]. Estimasi model yang kedua memiliki nilai AIC sebesar 56,97, sehingga model kedua dinilai lebih baik dalam melakukan klasifikasi dibandingkan model yang pertama.

Berikut adalah model persamaan logit terpilih yang terbentuk.

$$\ln(P(\widehat{Y} = 1)) = -3,53 + 3,44D_{anosmia} + 2,56D_{meriang} + 2,15D_{kontak} \tag{7}$$

Selanjutnya, pemeriksaan ketepatan model juga dilakukan melalui pembentukan confusion matrix for classification atau yang biasa disebut juga sebagai tabel kontingensi. Patro [18] mengungkapkan bahwa confusion matrix sangat berguna untuk mengetahui seberapa persen observasi dapat diklasifikasikan dengan tepat (akurasi).

Tabel 6. Confusion Matrix

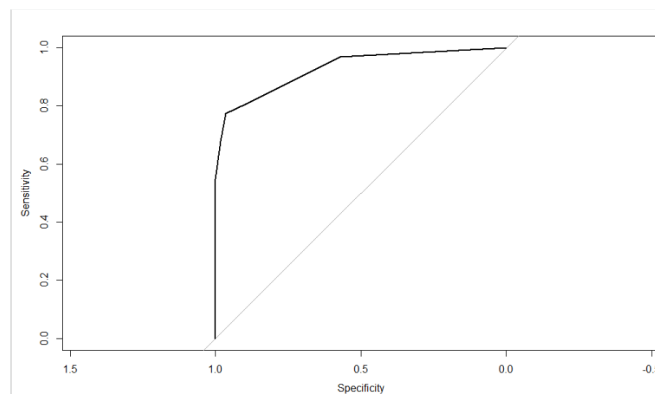
Observasi	Prediksi		Total
	Positif	Negatif	
Positif	24	2	26
Negatif	7	54	61
Total	31	56	87

$$Akurasi = \frac{True\ positive + True\ Negative}{Total} = \frac{24 + 54}{87} = 0,90 \tag{8}$$

Nilai ini dapat diartikan bahwa proporsi sampel yang dapat diklasifikasikan dengan tepat oleh model adalah 90%. Kemudian, dilakukan perhitungan sensitivitas dan spesifisitas model yang ada di persamaan (9) dan (10) untuk selanjutnya diplot ke dalam kurva ROC (Receiver Operator Characteristic) pada Gambar 5.

$$Sensitivitas = \frac{True\ positive}{True\ Positive + False\ Negative} = \frac{24}{24 + 2} = 0,92 \tag{9}$$

$$\text{Spesifisitas} = \frac{\text{True negative}}{\text{True negative} + \text{False positive}} = \frac{54}{54 + 7} = 0,89 \quad (10)$$



Gambar 5. Kurva ROC data *training*

Luas area di bawah kurva ROC pada data *training* adalah 0,93 sehingga diinterpretasi bahwa model dapat mengklasifikasikan data *training* dengan kategori sangat baik (*excellent*).

3.4. Interpretasi Variabel Prediktor pada Model Terbaik

3.4.1. Gejala Anosmia

Dari Tabel 5 diketahui bahwa *odds ratio* dari variabel dummy gejala anosmia adalah 31,29. Hal ini berarti bahwa kecenderungan seseorang yang mengalami anosmia untuk terinfeksi virus Covid-19 adalah sekitar 31 kalinya dibandingkan mereka yang tidak mengalami gejala tersebut. Tanda positif pada estimator parameter regresi logistik menunjukkan bahwa peluang seseorang untuk terinfeksi covid ketika mengalami gejala anosmia akan lebih besar. Sejalan dengan yang ditemukan oleh Oyelade, *et.al* [19] anosmia merupakan predictor paling kuat yang dilaporkan oleh penderita dibandingkan gejala demam, meskipun gejala demam adalah yang paling banyak dialami oleh pasien terkonfirmasi virus Covid-19. Selain itu, penelitian yang dilakukan oleh Elliot, *et.al* [9] juga menemukan hal yang sama, yaitu ketika test PCR hanya bisa dilaksanakan secara terbatas di suatu negara, keefektifan *screening* Covid-19 dapat dilakukan pada mereka yang mengalami 7 gejala utama, diantaranya anosmia dan meriang (*chill*).

Nilai *odds ratio* yang cukup besar memberikan implikasi seseorang yang telah mengalami gejala tersebut perlu diberikan peringatan dini untuk segera mendapatkan *treatment* sebagai pasien Covid-19. Jika memungkinkan untuk melakukan tes laboratorium maka sebaiknya segera dilakukan agar mendapatkan penanganan sesuai dengan tingkat gejalanya.

3.4.2. Gejala Meriang

Gejala meriang (*chill*) yang dicirikan oleh tubuh yang menggigil atau bahkan mengeluarkan keringat dingin. Pada model yang dihasilkan, terlihat bahwa koefisien regresi bertanda positif yang menandakan bahwa peluang seseorang untuk terkonfirmasi positif lebih besar ketika mengalami gejala ini. Nilai *odds ratio* sebesar

12,97 memiliki arti bahwa kecenderungan seseorang untuk terkonfirmasi positif adalah hampir 13 kalinya dibandingkan yang tidak mengalami gejala ini. Penemuan ini sejalan dengan temuan yang didapatkan Elliot, *et.al* [9] di Inggris Raya selama Bulan Juni 2020 hingga Januari 2021.

3.4.3. Kontak Erat dengan Pasien Positif Covid-19

Pada penelitian ini, kontak erat dengan pasien terinfeksi Covid-19 memiliki andil yang cukup besar dalam menentukan hasil tes. Seseorang yang pernah melakukan kontak dekat memiliki kecenderungan untuk mendapatkan hasil tes positif 8,62 atau hampir 9 kalinya dibandingkan yang merasa tidak pernah kontak. Temuan ini bersesuaian dengan hasil yang diperoleh oleh Song, *et.al* [16] di Huangzhou, China yang menunjukkan hubungan positif dan signifikan. Dengan demikian, penting sekali untuk memberikan *early warning* kepada seseorang yang pernah melakukan kontak erat dengan pasien untuk segera melakukan isolasi mandiri agar penyebaran virus dapat terkendali.

3.4.4. Validasi Model Menggunakan data Testing

Untuk mengetahui kekuatan model dalam melakukan prediksi terhadap data luar, dilakukan prediksi berdasarkan set data *testing*. Evaluasi ukuran ketepatan dilakukan dengan menghitung akurasi, sensitivitas, spesifisitas, dan mengukur luasan di bawah kurva ROC. Data testing yang digunakan sebanyak 20 data dari total sampel. *Confusion matrix* dari data *testing* disajikan pada Tabel 7.

Tabel 7. *Confusion Matrix* data testing

Observasi	Prediksi		Total
	Positif	Negatif	
Positif	4	1	5
Negatif	3	12	15
Total	7	13	20

$$Akurasi = \frac{True\ positive + True\ Negative}{Total} = \frac{4 + 12}{20} = 0,80 \quad (11)$$

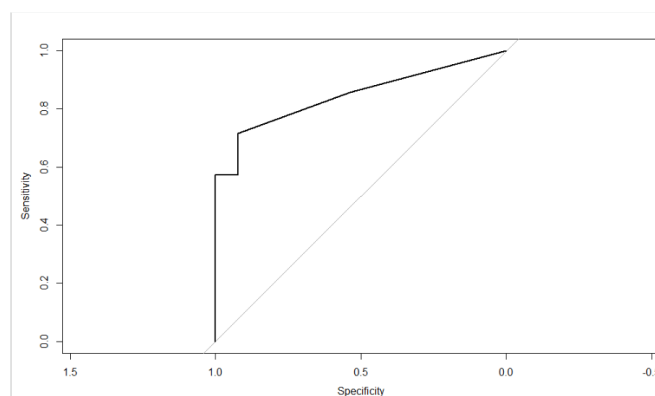
Akurasi model dalam mengklasifikasikan data *testing* sebesar 0,80 yang artinya, sebanyak 80 persen dari data testing dapat diklasifikasikan dengan baik oleh model yang terbentuk. Kemudian, perhitungan sensitivitas dan spesifisitas masing-masing menghasilkan nilai 0,80.

$$Sensitivitas = \frac{True\ positive}{True\ Positive + False\ Negative} = \frac{4}{4 + 1} = 0,80 \quad (12)$$

$$Spesifisitas = \frac{True\ negative}{True\ negative + False\ positive} = \frac{12}{12 + 3} = 0,80 \quad (13)$$

Dari persamaan (12) dan (13) kemudian dibuat kurva ROC pada Gambar 6.

Luas area di bawah kurva ROC di gambar 6 adalah 0,85 yang berarti bahwa kemampuan model dalam mengklasifikasikan sudah baik (*good*).



Gambar 6. Kurva ROC data testing

4. Kesimpulan

Penelitian menghasilkan model deteksi dini infeksi Covid-19 sebagai berikut:

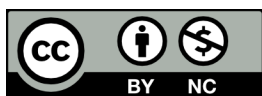
$$\ln(\widehat{P(Y = 1)}) = -3,53 + 3,44D_{anosmia} + 2,56D_{meriang} + 2,15D_{kontak}$$

Berdasarkan hasil estimasi dari model regresi logistik diketahui bahwa variabel yang berpengaruh pada penentuan klasifikasi seseorang terinfeksi virus Covid-19 adalah adalah gejala anosmia, gejala meriang, dan kontak erat. Oleh karena itu, seseorang yang paling tidak memiliki 1 dari ketiga kriteria tersebut, hendaknya mendapatkan treatment seperti pasien Covid-19. Model deteksi dini pada persamaan 8 sudah cukup baik dalam melakukan klasifikasi karena memiliki tingkat akurasi 0,80 spesifisifitas 0,80, dan sensitivitas 0,80, serta menghasilkan luas di bawah kurva ROC 0,85.

Referensi

- [1] D. A. D. Nasution, E. Erlina, and I. Muda, "Dampak Pandemi COVID-19 terhadap Perekonomian Indonesia," *Jurnal Benefita*, vol. 5, no. 2, pp. 212–224, jul 2020, doi: <http://dx.doi.org/10.22216/jbe.v5i2.5313>.
- [2] Anonim, "Covid-19," *Satuan Tugas Penanganan COVID-19*, 2020.
- [3] Komisi-IX, "Penurunan Harga Tes PCR Tak Selesaikan Masalah," *Dewan Perwakilan Rakyat Republik Indonesia*, 2021.
- [4] M. Qjidaa, Y. Mechbal, A. Ben-fares, H. Amakdouf, M. Maaroufi, B. Alami, and H. Qjidaa, "Early detection of COVID19 by deep learning transfer Model for populations in isolated rural areas," in *2020 International Conference on Intelligent Systems and Computer Vision (ISCV)*. IEEE, jun 2020, pp. 1–5, doi: <http://dx.doi.org/10.1109/ISCV49265.2020.9204099>.
- [5] A. Agustang, I. A. Mutiara, and A. Asrifan, "Genealogi Stigma Sosial Terhadap Pasien Covid 19," *OSF Preprint*, 2021, doi: <http://dx.doi.org/10.31219/osf.io/ep7jf>.
- [6] D. Junaedi, M. R. Arsyad, F. Salistia, and M. Romli, "Menguji Efektivitas Vaksinasi Covid-19 di Indonesia," *Reslaj : Religion Education Social Laa Roiba Journal*, vol. 4, no. 1, pp. 120–143, aug 2021, doi: <http://dx.doi.org/10.47467/reslaj.v4i1.537>.
- [7] S. Brief, "Smoking and COVID-19," *World Health Organization*, 2020.
- [8] A. Sultana, R. Sharma, M. M. Hossain, S. Bhattacharya, and N. Purohit, "Burnout among healthcare providers during COVID-19: Challenges and evidence-based interventions," *Indian Journal of Medical Ethics*, vol. 05, no. 04, pp. 308–311, nov 2020, doi: <http://dx.doi.org/10.20529/IJME.2020.73>.
- [9] J. Elliott, M. Whitaker, B. Bodinier, O. Eales, S. Riley, H. Ward, G. Cooke, A. Darzi, M. Chadeau-Hyam, and P. Elliott, "Predictive symptoms for COVID-19 in the community:

- REACT-1 study of over 1 million people," *PLOS Medicine*, vol. 18, no. 9, p. e1003777, sep 2021, doi: <http://dx.doi.org/10.1371/journal.pmed.1003777>.
- [10] D. M. N. Aditya, "Anosmia pada COVID-19: Studi Neurobiologi," *KELUWIH: Jurnal Kesehatan dan Kedokteran*, vol. 2, no. 1, pp. 50–55, dec 2020, doi: <http://dx.doi.org/10.24123/kesdok.V2i1.3098>.
- [11] Anonim, "Symptoms of COVID-19," *Centers for Disease Control and Prevention*, 2020.
- [12] —, "how is COVID-19 transmitted," *World Health Organization*, 2020.
- [13] L. Edi Nugroho and Arkham Zahri Rakhman, "Mobilitas Manusia dan Tingkat Penyebaran Covid-19: Sebuah Analisis Kuantitatif," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 10, no. 2, pp. 124–130, may 2021, doi: <http://dx.doi.org/10.22146/jnteti.v10i2.1519>.
- [14] M. Tirzite, M. Bukovskis, G. Strazda, N. Jurka, and I. Taivans, "Detection of lung cancer with electronic nose and logistic regression analysis," *Journal of Breath Research*, vol. 13, no. 1, p. 016006, nov 2018, doi: <http://dx.doi.org/10.1088/1752-7163/aae1b8>.
- [15] R. Xiao, X. Cui, H. Qiao, X. Zheng, and Y. Zhang, "Early diagnosis model of Alzheimer's Disease based on sparse logistic regression," *Multimedia Tools and Applications*, vol. 80, no. 3, pp. 3969–3980, jan 2021, doi: <http://dx.doi.org/10.1007/s11042-020-09738-0>.
- [16] C.-Y. Song, J. Xu, J.-Q. He, and Y.-Q. Lu, "COVID-19 early warning score: a multi-parameter screening tool to identify highly suspected patients," *medRxiv*, 2020.
- [17] A. Agresti, *Categorical Data Analysis*, 3rd ed. John Wiley & Sons Inc, 2018.
- [18] V. M. Patro and M. Ranjan Patra, "Augmenting Weighted Average with Confusion Matrix to Enhance Classification Accuracy," *Transactions on Machine Learning and Artificial Intelligence*, vol. 2, no. 4, pp. 77–91, aug 2014, doi: <http://dx.doi.org/10.14738/tmlai.24.328>.
- [19] O. N. Oyelade and A. E. Ezugwu, "A case-based reasoning framework for early detection and diagnosis of novel coronavirus," *Informatics in Medicine Unlocked*, vol. 20, p. 100395, 2020, doi: <http://dx.doi.org/10.1016/j.imu.2020.100395>.



This article is an open-access article distributed under the terms and conditions of the [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/). Editorial of JJoM: Department of Mathematics, Universitas Negeri Gorontalo, Jln. Prof. Dr. Ing. B.J. Habibie, Moutong, Tilongkabila, Kabupaten Bone Bolango, Provinsi Gorontalo 96119, Indonesia.