

Data Mining untuk Prediksi Status Pasien Covid-19 dengan Pengklasifikasi Naïve Bayes

Dewi Yanti Liliana¹, Hata Maulana², Agus Setiawan³

^{1,2,3}Jurusan Teknik Informatika dan Komputer, Politeknik Negeri Jakarta

Jl. Prof. G. A. Siwabessy, Kampus Baru Universitas Indonesia, 021-7270036

¹dewiyanti.liliana@tik.pnj.ac.id, ²hata.maulana@tik.pnj.ac.id, ³agus.setiawan@tik.pnj.ac.id

Diterima : 17 April 2021. Disetujui : 20 Juni 2021. Dipublikasikan : 22 Juni 2021.

Abstract - The Covid-19 pandemic in 2020 is a complex health problem and requires fast handling and collaborative solutions from various disciplines. Covid-19 patients who are hospitalized have different conditions and severity. This has an effect on the handling actions that will be taken by medical personnel. The large number of patients and the lack of medical personnel have resulted in the need for technology support to help classify patient status based on their conditions so that treatment is concentrated on patients who are very serious and need fast treatment. This study applies predictive techniques from data mining disciplines to classify the emergency status of patients. The Naive Bayes Classifier was applied to build a model based on a dataset of patients infected with Covid-19. The dataset of Covid-19 patients in Indonesia was obtained from www.kaggle.com and applied using RapidMiner software. The model built can predict the emergency status of patients based on age and sex who have the highest likelihood of recovering from Covid-19 and patients who have a high likelihood of continuing to undergo treatment and /or deceased. The results of this study indicates that the classification of the Naive Bayes reached 96.67% of accuracy rate in classifying patient status.

Keywords: Covid-19, naïve-bayes, classification, prediction, pandemic

Abstrak - Pandemi Covid-19 di tahun 2020 menjadi masalah kesehatan yang kompleks dan membutuhkan penanganan cepat serta kolaborasi solusi dari berbagai disiplin ilmu. Pasien Covid-19 yang mendapatkan perawatan di rumah sakit memiliki kondisi dan tingkat keparahan yang berbeda-beda. Hal ini berpengaruh pada tindakan penanganan yang akan dilakukan oleh petugas medis. Banyaknya pasien serta kurangnya tenaga medis mengakibatkan perlunya dukungan teknologi untuk membantu mengklasifikasikan status pasien berdasarkan kondisinya agar penanganan dikonsentrasikan pada pasien yang sangat gawat dan membutuhkan penanganan cepat. Penelitian ini menerapkan teknik prediksi dari disiplin ilmu data mining untuk mengklasifikasikan status kegawatan pasien. Pengklasifikasi Naive Bayes diterapkan untuk membangun model berdasarkan dataset pasien yang terinfeksi Covid-19. Dataset pasien Covid-19 di Indonesia diperoleh dari www.kaggle.com dan diaplikasikan menggunakan RapidMiner. Model yang dibangun dapat memprediksi status kegawatan pasien berdasarkan usia dan jenis kelamin yang memiliki kemungkinan tertinggi untuk sembuh dari Covid-19 dan pasien yang memiliki kemungkinan tinggi untuk tetap menjalani pengobatan dan atau meninggal dunia. Hasil penelitian ini menunjukkan bahwa klasifikasi metode Naive Bayes memiliki tingkat akurasi yang tinggi dalam mengklasifikasikan status pasien yaitu 96,67%.

Kata kunci: Covid-19, naïve-bayes, klasifikasi, prediksi, pandemi

I. PENDAHULUAN

Corona Virus Disease 2019 atau disingkat Covid-19 merupakan penyakit baru yang muncul di tahun 2019 dan dapat menyebabkan radang paru-paru dan gangguan pernapasan [1]. Penyakit ini disebabkan oleh Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) dan dapat menyebabkan kematian [2]. Dunia tengah dilanda pandemi Covid-19 yang melumpuhkan banyak sektor penting kehidupan baik di bidang ekonomi, sosial, dan keamanan [3]. Tatanan kehidupan dan tantangan akibat krisis Kesehatan global ini melanda semua negara di dunia. Kondisi yang sama juga terjadi di Indonesia [4]. Sejak awal maret, Pemerintah Indonesia telah mengumumkan jumlah

kasus Covid-19 yang dikonfirmasi, kasus yang pulih serta kasus kematian setiap harinya [5].

Pada bulan April 2021, angka kematian akibat virus Corona di Indonesia merupakan yang tertinggi di Asia setelah China, dengan korban meninggal 41.815 orang. Jumlah total kasus virus korona mencapai 1.537.967 kasus dengan 1.381.667 orang sembuh pada saat makalah ditulis [6]. Angka ini terus merangkak naik. Mengingat wabah Covid-19 merupakan masalah global di belahan dunia termasuk di Indonesia [7]. Studi ini dilakukan sebagai bagian dari upaya Mitigasi terhadap penyebaran penyakit Covid-19 di Indonesia.

Kebanyakan orang yang terinfeksi virus Covid-19 akan mengalami penyakit pernapasan ringan hingga sedang dan sembuh tanpa memerlukan perawatan [1]. Orang tua, dan mereka

yang memiliki masalah medis seperti penyakit kardiovaskular, diabetes, penyakit pernapasan kronis, dan kanker lebih mungkin untuk mengembangkan penyakit serius. Penyakit ini telah menjadi pandemi krisis kesehatan global sejak Maret 2020 [7] [8].

Banyaknya pasien yang di rumah sakit dan kapasitas tenaga medis menjadi permasalahan utama yang dihadapi di berbagai daerah [5]. Pasien dengan tingkat kegawatan tinggi memerlukan prioritas penanganan dibanding pasien dengan gejala sedang atau tanpa gejala [9]. Tenaga medis memerlukan bantuan untuk mengklasifikasi status pasien berdasarkan data pasien secara otomatis untuk mengurangi kelelahan tenaga medis yang harus terus bertugas dan meminimalisir resiko penanganan yang terlambat terhadap pasien [10]. Oleh karena itu dibutuhkan solusi teknologi berbasis data secara otomatis yang dapat membantu mengklasifikasikan status kegawatan berdasarkan data pasien.

Data mining adalah teknik yang digunakan untuk membangun model pembelajaran mesin. Pembelajaran mesin (*machine learning*) adalah teknik kecerdasan buatan modern yang belajar membangun model dengan menggunakan data empiris [11]. Data Mining digunakan untuk menemukan pola dalam kumpulan besar data mentah. Data Mining menerapkan teknik *Machine Learning* untuk menarik pengetahuan pada data. Dalam penelitian ini penulis menerapkan teknik data mining untuk mengklasifikasikan dataset Covid-19 menggunakan Algoritma *Naive Bayes Classifier* (NBC) karena NBC telah berhasil diterapkan dalam banyak tugas klasifikasi berbasis probabilitas bersyarat pada populasi data [12].

Penelitian ini bertujuan memberikan solusi untuk mengklasifikasikan status pasien Covid-19 secara otomatis berdasarkan gejala yang dialami pasien. Klasifikasi dilakukan menggunakan algoritma Naïve-Bayes yang dikenal memiliki akurasi tinggi dengan pembelajaran tersupervisi berbasis distribusi dataset. Pemilihan algoritma yang tepat diharapkan dapat menghasilkan model klasifikasi yang baik dan berguna untuk diimplementasikan secara riil pada penanganan dan skrining status kegawatan pasien Covid-19.

Kontribusi utama pada penelitian ini adalah menghasilkan model prediktif status pasien Covid-19 sehingga dapat disiapkan langkah-langkah yang cepat dan tepat untuk penanganan pasien. Algoritma NBC yang memiliki akurasi tinggi dan telah banyak diterapkan pada bermacam permasalahan prediksi [13], sehingga diharapkan dapat membantu memprediksi status pasien untuk mendapatkan penanganan yang tepat secara langsung oleh tim medis.

II. METODE PENELITIAN

Penelitian ini memprediksi status kegawatan pasien menggunakan teknik data mining berbasis dataset pasien. Metode yang digunakan dalam penelitian ini dimulai dengan: 1) kaji literatur; 2) pengumpulan dataset pasien; 3) pembangunan model Naïve Bayes Classifier; 4) Uji coba dan validasi hasil. Penjelasan setiap langkah dalam metode penelitian dijelaskan sebagai berikut.

A. Kaji Literatur

Pengkajian literatur sangat penting dilakukan untuk mengetahui penerapan metode untuk menyelesaikan problem yaitu klasifikasi status pasien. Literatur utama, yaitu *data mining*. Penambangan data atau *data mining* adalah proses yang digunakan untuk mengubah data mentah menjadi informasi yang berguna [14]. Dengan bantuan perangkat lunak spesifik, data mining dapat digunakan untuk mencari pola dalam kumpulan data yang besar, sebuah bisnis dapat mempelajari lebih lanjut tentang pelanggan mereka untuk mengembangkan strategi pemasaran yang lebih efektif, meningkatkan penjualan, dan menurunkan biaya. Penambangan data bergantung pada pengumpulan data yang efektif, data warehousing, dan pemrosesan komputer [15].

Pengklasifikasi Naïve Bayes atau NBC didasarkan pada teorema Bayes dengan asumsi independensi diantara variabel prediktor. Model Naive Bayes mudah dibuat dan sangat berguna untuk prediksi pada kumpulan data yang sangat besar [12]. Selain sederhana dan mudah dipahami, NBC dikenal memiliki kinerja yang bahkan mengungguli metode klasifikasi yang sangat canggih [16]. Teorema Bayes digambarkan pada gambar 1.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Gambar 1. Penjabaran Persamaan Teorema Bayes

Pada gambar 1, persamaan probabilitas Bayes memiliki komponen diantaranya yaitu:

- A : Hipotesis data berada pada suatu kelas tertentu
- B : Data dengan kelas yang belum diketahui
- $P(A|B)$: Hipotesis probabilitas berdasarkan syarat/kondisi tertentu
- $P(A)$: Probabilitas dari suatu hipotesis kelas A
- $P(B|A)$: Probabilitas berdasarkan kondisi hipotesis kelas A
- $P(B)$: Probabilitas dari data yang tersedia

Secara ekivalen, Teorema Bayes dapat ditulis dengan persamaan 1.

$$Posterior = \frac{Likelihood * Prior}{Evidence} \quad (1)$$

Dengan *Posterior* sebagai nilai probabilitas data yang akan diprediksi kelasnya, *likelihood* adalah probabilitas atribut-atribut pada data dengan kondisi/syarat A, prior adalah probabilitas kelas A dalam populasi, dan *evidence* adalah probabilitas karakteristik sampel secara global.

B. Pengumpulan Dataset

Dataset pasien Covid-19 di Indonesia diperoleh dari website Kaggle (www.kaggle.com) [17]. Dataset inilah yang nantinya akan dilatih untuk menghasilkan model prediksi terhadap status pasien. Dataset disiapkan dan dibersihkan di mana hanya atribut relevan yang diekstraksi dan digunakan dalam proses klasifikasi. Dataset pasien terdiri dari 156 baris data dengan empat atribut. Dataset pasien digambarkan pada gambar 2.

gender	age	nationality	province	current state
female	30-40	indonesia	DKI Jakarta	released
female	60-70	indonesia	DKI Jakarta	isolated
female	30-40	indonesia	DKI Jakarta	released
female	30-40	indonesia	DKI Jakarta	isolated
male	50-60	indonesia	DKI Jakarta	isolated
male	30-40	indonesia	DKI Jakarta	released
female	50-60	indonesia	DKI Jakarta	released
male	50-60	indonesia	DKI Jakarta	released
female	50-60	indonesia	DKI Jakarta	released
male	20-30	foreigner	DKI Jakarta	isolated
female	50-60	foreigner	Yogyakarta	isolated
male	30-40	indonesia	DKI Jakarta	isolated
female	20-30	indonesia	DKI Jakarta	released
male	50-60	indonesia	DKI Jakarta	isolated
female	40-50	indonesia	Yogyakarta	released
female	20-30	indonesia	Jawa Barat	isolated
male	50-60	indonesia	DKI Jakarta	isolated
male	50-60	indonesia	Banten	isolated

Gambar 2. Dataset Pasien dari Website Kaggle (www.kaggle.com)

Atribut yang digunakan adalah jenis kelamin, usia, kewarganegaraan, dan provinsi. Sedangkan target status pasien ada tiga kategori yaitu ringan, sedang, dan berat (*released, isolated, deceased*).

C. Pembangunan Model Naïve Bayes Classifier

Naive Bayes Classifier adalah salah satu algoritma klasifikasi data mining dan digunakan untuk membedakan setiap data pada dataset berdasarkan atribut yang ditentukan. Langkah pembangunan model NBC adalah sebagai berikut:

1. Perhitungan Probabilitas Prior (P(A_i))

Menghitung probabilitas prior dari setiap kategori kelas.

1. A₀ (Status = “isolated”)

Probabilitas pasien berstatus *isolated*

$$P(A_0) = 131/156 = 0.84$$

2. A₁ (Status = “released”)

Probabilitas pasien berstatus *released*

$$P(A_1) = 9/156 = 0.058$$

3. A₃ (Status = “deceased”)

Probabilitas pasien berstatus *deceased*

$$P(A_3) = 16/156 = 0.10$$

2. Perhitungan Probabilitas Likelihood (P(B|A_i))

Probabilitas *likelihood* dapat dihitung dengan terlebih dahulu membuat tabel frekuensi untuk setiap atribut terhadap target kelas. Perhitungan probabilitas *likelihood* dilakukan pada 156 data latih dengan menggunakan *instance* data B sebagai vektor. Pemilihan atribut klasifikasi pasien adalah jenis kelamin, usia, kewarganegaraan, provinsi. Gambar 3 hingga gambar 6 menunjukkan perhitungan probabilitas atribut jenis kelamin, usia, provinsi, dan kebangsaan., Naive Bayes adalah salah satu algoritma klasifikasi data mining dan digunakan untuk membedakan *instance* dataset berdasarkan atribut yang telah ditentukan.

GENDER	ISOLATED	RELEASED	DECEASED	P(ISOLATED)	P(RELEASED)	P(DECEASED)
MALE	74	4	13	0,582677165	0,333333333	0,764705882
FEMALE	53	8	4	0,417322835	0,666666667	0,235294118
Total	127	12	17			

Gambar 3. Probabilitas Jenis Kelamin

AGE	ISOLATED	RELEASED	DECEASED	P(ISOLATED)	P(RELEASED)	P(DECEASED)
0-10	1	1	1	0,007874016	0,083333333	0,058823529
20-30	20	1	0	0,157480315	0,083333333	0
30-40	21	3	2	0,165354331	0,25	0,117647059
40-50	27	1	2	0,212598425	0,083333333	0,117647059
50-60	29	4	7	0,228346457	0,333333333	0,411764706
60-70	15	1	3	0,118110236	0,083333333	0,176470588
70-80	8	1	2	0,062992126	0,083333333	0,117647059
80-90	6	0	0	0,047244094	0	0
Total	127	12	17			

Gambar 4. Probabilitas Usia

PROVINCE	ISOLATED	RELEASED	DECEASED	P(ISOLATED)	P(RELEASED)	P(DECEASED)
BALI	1	1	1	0,007874016	0,083333333	0,058823529
BANTEN	6	0	1	0,047244094	0	0,058823529
DKI JAKARTA	98	9	11	0,771653543	0,75	0,647058824
JAWA BARAT	8	0	1	0,062992126	0	0,058823529
JAWA TENGAH	3	0	2	0,023622047	0	0,117647059
JAWA TIMUR	7	0	0	0,055118111	0	0
KALIMANTAN BARAT	2	0	0	0,015748031	0	0
KEPULAUAN RIAU	0	1	0	0	0,083333333	0
SULAWESI UTARA	1	0	0	0,007874016	0	0
YOGYAKARTA	1	1	1	0,007874016	0,083333333	0,058823529
Total	127	12	17			

Gambar 5. Probabilitas Provinsi

NATIONALITY	ISOLATED	RELEASED	DECEASED	P(ISOLATED)	P(RELEASED)	P(DECEASED)
INDONESIA	119	11	17	0,937007874	0,916666667	1
FOREIGNER	8	1	0	0,062992126	0,083333333	0
Total	127	12	17			

Gambar 6. Probabilitas Kewarganegaraan

3. Perhitungan Probabilitas Posterior P(A|B_i)

Perhitungan probabilitas posterior dilakukan untuk memprediksi peluang suatu data tes merupakan anggota kelas A. Perhitungan pada suatu data uji x_1 dan x_2 (gambar 7) menggunakan algoritma NBC dijelaskan sebagai berikut.

gender	age	nationality	province	current_state
male	30-40	indonesia	DKI Jakarta	?
female	60-70	indonesia	DKI Jakarta	?

Gambar 7. Contoh Data Uji

Berdasarkan pengujian data pertama (x_1) pada gambar 7, maka dapat dilakukan klasifikasi status pasien Covid-19 dengan ketentuan nilai probabilitas masing-masing atribut dikalikan untuk mendapatkan nilai probabilitas *posterior*. Perhitungan *evidence* dapat diabaikan dengan asumsi nilai *evidence* selalu konsisten untuk semua kelas A_i . Perhitungan nilai probabilitas *likelihood* pada suatu data test x_i dengan persamaan 2.

$$P(x_i|A_n) = \prod_{j=0}^m P(x_{ij}|A_n) \quad (2)$$

Dimana $P(x_i|A_n)$ adalah probabilitas *likelihood* untuk banyak fitur independen; x_i adalah data uji ke- i ; A_n adalah target kelas yang terdiri dari $1, \dots, n$ kelas; dan x_{ij} adalah data uji dengan fitur/atribut ke $j=1, \dots, m$. Dalam kasus klasifikasi pasien ini, empat atribut berturut-turut adalah jenis kelamin, usia, provinsi, dan kewarganegaraan. Nilai *likelihood* setiap atribut dapat dilihat pada gambar 3 sampai dengan gambar 6.

Perhitungan probabilitas *likelihood* untuk data uji pertama pada gambar 7 terhadap status kelas A_0 (*isolated*) adalah sebagai berikut:

$$\begin{aligned} P(x_1|A_0) &= P(x_{11}|A_0) * P(x_{12}|A_0) * P(x_{13}|A_0) * P(x_{14}|A_0) \\ P(x_1|A_0) &= 0.583 * 0.165 * 0.937 * 0.771 \\ &= 0.070 \end{aligned}$$

Perhitungan probabilitas *likelihood* status kelas A_1 (*Released*) adalah sebagai berikut:

$$\begin{aligned} P(x_1|A_1) &= P(x_{11}|A_1) * P(x_{12}|A_1) * P(x_{13}|A_1) * P(x_{14}|A_1) \\ &= 0.333 * 0.250 * 0.917 * 0.750 \\ &= 0.057 \end{aligned}$$

Perhitungan probabilitas *likelihood* status kelas A_3 (*Deceased*) adalah sebagai berikut:

$$\begin{aligned} P(x_1|A_2) &= P(x_{11}|A_2) * P(x_{12}|A_2) * P(x_{13}|A_2) * P(x_{14}|A_2) \\ &= 0.765 * 0.118 * 1 * 0.647 \\ &= 0.058 \end{aligned}$$

Selanjutnya dilakukan perhitungan probabilitas *posterior* dengan persamaan 3.

$$P(A_n|x_i) = P(x_i|A_n) * P(A_n) \quad (3)$$

Dimana $P(A_n|x_i)$ adalah probabilitas *posterior* kelas $1, \dots, n$ dan $P(A_n)$ adalah probabilitas *prior*. Sehingga dapat dihitung probabilitas *posterior* masing-masing kelas status pasien.

Probabilitas *posterior* status pasien x_1 *isolated*:

$$\begin{aligned} P(A_0|x_1) &= P(x_1|A_0) * P(A_0) \\ &= 0.07 * 0.84 \\ &= 0.0588 \end{aligned}$$

Probabilitas *posterior* status pasien x_1 *released*:

$$\begin{aligned} P(A_1|x_1) &= P(x_1|A_1) * P(A_1) \\ &= 0.057 * 0.058 \\ &= 0.0033 \end{aligned}$$

Probabilitas *posterior* status pasien x_1 *deceased*:

$$\begin{aligned} P(A_2|x_1) &= P(x_1|A_2) * P(A_2) \\ &= 0.058 * 0.103 \\ &= 0.006 \end{aligned}$$

Prediksi NBC dilakukan dengan memilih probabilitas *posterior* maksimum dari ketiga kelas. Hasil prediksi data uji pasien pertama adalah berstatus *isolated* dengan nilai $P(A_0|x_1) > P(A_1|x_1) > P(A_2|x_1)$.

Proses perhitungan yang sama juga dilakukan terhadap data uji kedua yang memberikan hasil probabilitas *posterior* sebagai berikut.

Probabilitas *posterior* status pasien x_2 *isolated*:

$$\begin{aligned} P(A_0|x_2) &= P(x_2|A_0) * P(A_0) \\ &= 0.036 * 0.84 \\ &= 0.03 \end{aligned}$$

Probabilitas *posterior* status pasien x_2 *released*:

$$\begin{aligned} P(A_1|x_2) &= P(x_2|A_1) * P(A_1) \\ &= 0.382 * 0.058 \\ &= 0.022 \end{aligned}$$

Probabilitas *posterior* status pasien x_2 *deceased*:

$$\begin{aligned} P(A_2|x_2) &= P(x_2|A_2) * P(A_2) \\ &= 0.027 * 0.103 \\ &= 0.003 \end{aligned}$$

Berdasarkan perhitungan probabilitas *posterior* data uji pasien kedua, maka hasil prediksi status pasien adalah *isolated* karena nilai $P(A_0|x_2) > P(A_1|x_2) > P(A_2|x_2)$.

III. HASIL DAN PEMBAHASAN

Implementasi data mining dilakukan menggunakan perangkat lunak *RapidMiner*. Dataset untuk pelatihan dan pengujian model NBC diperoleh dari data pasien Indonesia-Coronavirus yang tersedia di website Kaggle.com. Atribut yang terdapat dalam kumpulan data pasien adalah jenis kelamin, usia, kewarganegaraan, provinsi, dan negara bagian serta target kelas yaitu status pasien. Status pasien ada tiga yaitu *isolated*, *released*, dan *deceased*.

Gambar 8 menunjukkan desain solusi RapidMiner untuk klasifikasi data pasien menggunakan *Naive Bayes Classifier*.

Gambar 8. Implementasi NBC dengan RapidMiner Studio

Tujuan utama dari penelitian ini adalah untuk memprediksi nilai probabilitas dengan algoritma *Naive Bayes* yang digunakan untuk mengklasifikasikan status pasien Covid-19. Pada Desain RapidMiner di Gambar 8, pada tahap pelatihan terdapat operator *Read Excel* untuk memilih dataset dan operator *Naive Bayes* untuk mengklasifikasi dataset. Sedangkan pada pengujian terdapat operator *Apply Model* untuk menjalankan model naive bayes dan operator *Performance* untuk mengukur prediksi performa model NBC.

Hasil prediksi klasifikasi model Naive Bayes menggunakan perangkat lunak *RapidMiner* konsisten dengan perhitungan manual yang telah dilakukan pada bab 2, ditunjukkan pada Gambar 9.

Row No.	current_state	prediction(c...	confidence(...	confidence(...	confidence(...	gender	age
1	?	isolated	0.065	0.841	0.094	male	30-40
2	?	isolated	0.084	0.832	0.084	female	60-70

Gambar 9. Hasil Prediksi dengan *RapidMiner*

Hasil pengujian terhadap dua contoh data uji Gambar 7 menunjukkan bahwa kedua pasien diprediksi berstatus *isolated*. Hal ini juga sesuai dengan label asli data uji kedua pasien yang berstatus *isolated*.

Selanjutnya untuk menghitung performa klasifikasi dilakukan pengujian terhadap 30 data uji yang terdiri dari 26 data berstatus *isolated*, 3 data berstatus *released*, dan 1 data berstatus *deceased*. Data uji yang digunakan diluar data latih yang berjumlah 156 data. Hasil pengujian dinyatakan dalam nilai presisi, *recall*, dan akurasi. Pada *RapidMiner* terdapat operator untuk menampilkan hasil performa model yang dinyatakan dalam bentuk *confussion matrix*. Gambar 10 menunjukkan *confussion matrix* hasil pengujian.

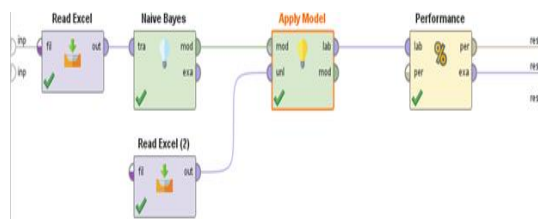
	true isolated	true released	true deceased	precision
pred. isolated	25	0	0	100%
pred. released	1	3	0	75%
pred. deceased	0	0	1	100%
recall	96,15%	100%	100%	

Gambar 10. *Confussion Matrix* Hasil Pengujian

Confussion matrix pada Gambar 10 menunjukkan beberapa informasi penting, diantaranya yaitu jumlah data uji riil dan jumlah data uji hasil prediksi untuk setiap kelasnya, presisi, dan *recall* setiap kelas target. *Confussion matrix* juga menginformasikan banyaknya nilai *true positive*,

true negative, *false positive*, dan *false negative* dari pengujian.

Nilai presisi menunjukkan seberapa tepat



prediksi dalam mengenali data sesuai kelas aslinya, sedangkan nilai *recall* menunjukkan kemampuan *classifier* dalam menemukan kembali informasi sesuai kelas aslinya. Akurasi menunjukkan performa keseluruhan dari model yang dibangun. Hasil pengujian dalam bentuk *confussion matrix* tiga kelas status pasien dapat dianalisis dalam bentuk rata-rata presisi, rata-rata *recall* dan akurasi yang ditunjukkan pada Tabel 1.

TABEL 1. PERFORMA HASIL PENGUJIAN

Parameter	Nilai
Presisi	92%
<i>Recall</i>	98,72%
Akurasi	96,67%

Berdasarkan Tabel 1 maka dapat dilihat bahwa akurasi klasifikasi dengan NBC mencapai 96.67%. Model klasifikasi memiliki rata-rata *recall* yang lebih tinggi (98.72%) dari rata-rata presisi (92%). Hal ini berarti bahwa model yang dibangun memiliki sensitivitas yang baik dalam mengenali kelas data uji yang ada, sedangkan presisi meyakinkan ketepatan hasil prediksi. Presisi dan *recall* Bersama-sama digunakan untuk menunjukkan seberapa baik prediksi suatu model pengklasifikasi.

IV. KESIMPULAN

Pada penelitian ini dikembangkan model *Data Mining* untuk prediksi status pasien Covid-19 menggunakan algoritma *Naive Bayes Classifier* atau NBC. NBC digunakan karena kehandalannya dalam mengklasifikasi data berdasarkan atribut-atribut yang dimiliki, baik berupa nilai numerik maupun kategorik. Model dibangun menggunakan dataset pasien Covid-19 yang didapat dari website www.kaggle.com dan diimplementasikan menggunakan perangkat lunak *RapidMiner*. Hasil dari model NBC untuk prediksi status pasien Covid-19 di Indonesia memberikan hasil yang diukur dalam nilai presisi, *recall*, dan akurasi, berturut-turut nilainya 92%, 88.72%, dan 96.67%. Nilai akurasi yang tinggi menunjukkan kinerja NBC yang baik

dalam mengklasifikasikan tiga status pasien, yaitu *isolated*, *released*, dan *deceased*. Hasil dari penelitian ini bermanfaat untuk diterapkan pada situasi nyata, untuk membantu tenaga medis menentukan tindakan. Ke depan, jumlah dataset nyata dan berukuran besar dengan proporsi nilai setiap kelas yang seimbang sangat baik untuk mendapatkan akurasi prediksi yang lebih tinggi.

REFERENSI

- [1] C. Long *et al.*, "Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT?," *Eur. J. Radiol.*, vol. 126, p. 108961, May 2020, doi: 10.1016/j.ejrad.2020.108961.
- [2] S. Sanche, Y. T. Lin, C. Xu, E. Romero-Severson, N. Hengartner, and R. Ke, "High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2," *Emerg. Infect. Dis. J.*, vol. 26, no. 7, 2020, doi: 10.3201/eid2607.200282.
- [3] F. Rustam *et al.*, "COVID-19 Future Forecasting Using Supervised Machine Learning Models," *IEEE Access*, vol. 8, pp. 101489–101499, 2020, doi: 10.1109/ACCESS.2020.2997311.
- [4] D. Susanna, "When will the COVID-19 pandemic in Indonesia end?," *Kesmas*, vol. 15, no. 4, pp. 160–162, 2020, doi: 10.21109/KESMAS.V15I4.4361.
- [5] M. Sukmana, M. Aminuddin, and D. Nopriyanto, "Indonesian government response in COVID-19 disaster prevention," *East African Sch. J. Med. Sci.*, vol. 3, no. 3, pp. 81–6, 2020, doi: 10.36349/EASMS.2020.v03i03.025.
- [6] "KawalCovid." [Online]. Available: <https://kawalcovid19.id/>.
- [7] S. K. Kar, S. M. Y. Arafat, P. Sharma, A. Dixit, M. Marthoenis, and R. Kabir, "COVID-19 pandemic and addiction: Current problems and future concerns," *Asian J. Psychiatr.*, vol. 51, p. 102064, 2020, doi: <https://doi.org/10.1016/j.ajp.2020.102064>.
- [8] Y. C. Wu, C. S. Chen, and Y. J. Chan, "The outbreak of COVID-19: An overview," *J. Chinese Med. Assoc.*, vol. 83, no. 3, pp. 217–220, 2020, doi: 10.1097/JCMA.0000000000000270.
- [9] M. Abed Alah, S. Abdeen, and V. Kehyayan, "The first few cases and fatalities of Corona Virus Disease 2019 (COVID-19) in the Eastern Mediterranean Region of the World Health Organization: A rapid review," *J. Infect. Public Health*, vol. 13, no. 10, pp. 1367–1372, 2020, doi: 10.1016/j.jiph.2020.06.009.
- [10] L. K. Kumar and P. J. A. Alphonse, "Automatic Diagnosis of COVID-19 Disease using Deep Convolutional Neural Network with Multi-Feature Channel from Respiratory Sound Data : Cough , Voice , and Breath Reference : To appear in : Received Date : Revised Date : Accepted Date : Abstract ;," *Alexandria Eng. J.*, 2021, doi: 10.1016/j.aej.2021.06.024.
- [11] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artif. Intell.*, vol. 267, pp. 1–38, 2019, doi: 10.1016/j.artint.2018.07.007.
- [12] A. Fattah and R. Setyadi, "Teknologi informasi dan pendidikan," *J. Teknol. Inf. dan Pendidik.*, vol. 12, no. 2, pp. 1–7, 2019.
- [13] P. D. Utami and R. Sari, "Filtering Hoax Menggunakan Naive Bayes Classifier," *Multinetics*, vol. 4, no. 1, p. 57, 2018, doi: 10.32722/vol4.no1.2018.pp57-61.
- [14] Prasdika and B. Sugiantoro, "A Review Paper On Big Data And Data Mining," *Int. J. Informatics Dev.*, vol. 7, no. 1, pp. 33–35, 2018.
- [15] M. H. Frické, "Data-Information-Knowledge-Wisdom (DIKW) Pyramid, Framework, Continuum," in *Encyclopedia of Big Data*, L. A. Schintler and C. L. McNeely, Eds. Cham: Springer International Publishing, 2018, pp. 1–4.
- [16] D. Y. Liliana and D. Priharsari, "Tsunami Early Warning Detection using Bayesian Classifier," *Proc. - 2019 2nd Int. Conf. Comput. Informatics Eng. Artif. Intell. Roles Ind. Revolut. 4.0, IC2IE 2019*, pp. 44–48, 2019, doi: 10.1109/IC2IE47452.2019.8940823.
- [17] "Kaggle." [Online]. Available: www.kaggle.com.