

# PREDIKSI KESEMBUHAN PASIEN COVID-19 DI INDONESIA MELALUI TERAPI MENGGUNAKAN METODE NAÏVE BAYES

Okky Putra Barus<sup>1\*)</sup>, Anton Tehja<sup>2)</sup>

<sup>1</sup>Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pelita Harapan  
E-mail : okky.barus@uph.edu<sup>1)</sup>

<sup>2</sup>Informatika, Fakultas Ilmu Komputer, Universitas Pelita Harapan  
E-mail : at70004@student.uph.edu<sup>2)</sup>

*\*Penulis Korespondensi*

---

*Abstract – This study aims to predict the recovery of COVID-19 patients in Indonesia by using Data Mining calculations. The method used to predict the recovery of COVID-19 patients is the Naïve Bayes method. The collection of datasets through trusted sources, the NIHR Innovation Observatory and datasets on an international/global scale, totaling 367 pieces of raw data that have not been filtered. After conducting the data feasibility test, the remaining 286 pieces of data will be divided into 70% of training data of 200 pieces of data and 30% of testing data of 86 pieces of data. Based on the test results, the use of the Naïve Bayes method in predicting the recovery of COVID-19 patients obtained an Accuracy of 96.51%, a Success Precision (Yes) of 100% and a Failure (No) of 95.71%, and a Success Sensitivity (Yes) of 84.21% and Failed (No) by 100%. Therefore, it is concluded that calculations using the Naïve Bayes method in this study will produce an accuracy rate of COVID-19 recovery of 96.51%, which means that the results of the predictions' calculation of success and failure in a therapy given to patients can be accounted for as data reference in a more detailed subsequent research..*

**Keywords:** COVID-19, data mining, Naïve Bayes, prediction

*Abstrak – Penelitian ini bertujuan untuk memprediksi kesembuhan pasien COVID-19 di Indonesia dengan menggunakan perhitungan Data Mining. Metode yang digunakan untuk memprediksi kesembuhan pasien COVID-19 adalah metode Naïve Bayes. Pengumpulan dataset melalui sumber yang terpercaya, NIHR Innovation Observatory dan dataset bersifat skala internasional/global yang berjumlah 367 buah data mentah yang belum dilakukan penyaringan. Setelah melakukan uji kelayakan data, yang tersisa adalah 286 buah data yang akan dibagi menjadi 70% data training sebesar 200 buah data dan 30% data testing sebesar 86 buah data. Berdasarkan hasil pengujian, penggunaan metode Naïve Bayes dalam memprediksi kesembuhan pasien COVID-19 mendapatkan Akurasi sebesar 96,51%, Presisi Sukses (Yes) sebesar 100% dan Gagal (No) sebesar 95,71%, dan Sensitifitas Sukses (Yes) sebesar 84,21% dan Gagal (No) sebesar 100%. Maka dari itu, diperoleh kesimpulan perhitungan menggunakan metode Naïve Bayes pada penelitian ini akan menghasilkan tingkat keakuratan kesembuhan COVID-19 sebesar 96,51% yang artinya hasil perhitungan Diterima <22062021>, Revisi <24062021>, Diterima untuk publikasi <26072021>. Copyright © 2021 Published by PSDKU Sistem Informasi UPH Kampus Medan, ISSN : 2528-5114*

*prediksi sukses dan gagalnya dalam suatu terapi yang diberikan kepada pasien dapat dipertanggungjawabkan sebagai acuan data dalam penelitian berikutnya yang lebih mendetail.*

**Kata Kunci:** COVID-19, data mining, Naïve Bayes, prediksi

## PENDAHULUAN

Pneumonia yang belum diketahui secara pasti oleh ilmuwan penyebabnya terdeteksi pertama kali di Wuhan, Cina, dilaporkan ke *World Health Organization* (WHO) di Cina pada 31 Desember 2019. Sejak itu, jumlah kasus virus Corona semakin meningkat seiring dengan tingginya angka kematian. Virus Corona menyebar dari satu kota ke seluruh negeri hanya dalam waktu 30 hari, pada 11 Februari 2020 lalu dinamai COVID-19 oleh WHO.

Selama satu tahun terakhir, virus Corona telah menyebar dengan cepat ke seluruh belahan dunia. Para ilmuwan terus mengembangkan obat untuk melawan penyakit ini. Karena jumlah kasus yang diuji untuk virus Corona meningkat dengan cepat dari hari ke hari, tidak mungkin untuk melakukan pengujian karena faktor waktu dan biaya. Selama beberapa tahun terakhir, *machine learning* dan *data mining* telah menjadi sangat berguna hingga menjadi solusi di bidang medis. Menggunakan teknik *machine learning* dan *data mining* untuk memprediksi COVID-19 pada pasien akan mempersingkat waktu tunda hasil tes medis dan mengorganisir petugas kesehatan untuk memberikan perawatan medis yang tepat kepada mereka.

*Data mining* memainkan peran penting dalam mencegah penyebaran virus ini. Karena peran ahli epidemiologi perawatan kesehatan telah berkembang, data kesehatan berbentuk *soft-copy* (elektronik) yang tersebar juga semakin meluas. Peningkatan ketersediaan data kesehatan elektronik memberikan peluang besar dalam perawatan kesehatan baik untuk penemuan maupun

pengaplikasian untuk meningkatkan perawatan kesehatan. Data ini dapat digunakan untuk melatih algoritma *data mining* guna menghitung prediksi tingkat kesembuhan pasien COVID-19 di Indonesia dalam hal memprediksi kasus penyakit yang akan datang.[1]

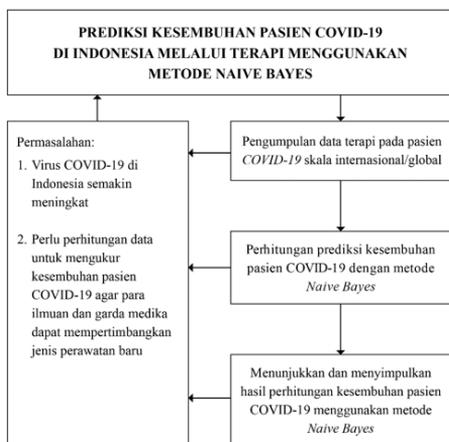
Tingkat kesembuhan pada seorang pasien COVID-19 sangat dipengaruhi banyak oleh beberapa faktor, seperti faktor lingkungan, *antibody*, nutrisi, kegiatan sehari-hari, obat yang dikonsumsi, frekuensi olahraga, dll. Faktor yang akan dibahas pada penelitian ini adalah faktor terapi obat yang diaplikasikan kepada pasien COVID-19, maka dari itu pada penelitian ini lebih difokuskan pada terapi penyembuhan melalui obat-obatan, untuk faktor-faktor lain yang mempengaruhi tidak akan dibahas pada penelitian ini.

Jenis terapi yang diberikan kepada pasien pada dataset penelitian ini terdapat dua yaitu Monoterapi dan Kombinasi (*Monotherapy and Combination*). Dimana Monoterapi, berarti hanya pasien yang diobati dengan obat antibiotik tunggal (satu jenis saja) yang diberikan, sedangkan pada kelompok terapi Kombinasi, pasien yang diobati dengan dua atau lebih jenis antibiotik yang diberikan secara bersamaan pada satu waktu.[2]

Hingga 8 Juni 2021, sebesar 173,694,649 kasus COVID-19 telah terdaftar secara skala global dan sebesar 1,869,325 kasus telah terdaftar di Indonesia beserta jumlah kematian mencapai 3,739,274 jiwa secara skala global dan sebanyak 51,992 jiwa di Indonesia dan jumlah pasien yang telah pulih sebesar 1,377,378,191 jiwa secara skala global

dan sebesar 22,536,306 jiwa di Indonesia [5]. Maka dari itu penelitian ini akan berfokus untuk menemukan terapi terbaik untuk menyembuhkan pasien COVID-19 berdasarkan data yang tersedia sehingga diharapkan dapat memberikan rekomendasi terapi jenis apa yang terbaik bagi pasien COVID-19.

**METODE PENELITIAN**



Penelitian ini dimulai dari pengumpulan data secara mendasar menggunakan metode studi Pustaka (dari Internet) untuk kasus COVID-19 di berbagai belahan dunia. Sumber yang didapatkan melalui situs *NIHR Innovation Observatory*, dimana *NIHR Innovation Observatory* merupakan fasilitas pemindaian medis nasional yang berlokasi di Universitas Newcastle[4]. Hasil data yang dikumpulkan merupakan dataset yang berisi jenis terapi / *treatment* apa saja yang telah diujikan kepada pasien pengidap virus COVID-19 dan hasil terapi / *treatment* nya berhasil atau tidak.

Dari data-data yang sudah terkumpul, maka dapat digunakan untuk perhitungan berikutnya untuk memprediksi apakah seorang pasien dapat sembuh menggunakan terapi / *treatment* tertentu. Penjelasan mengenai variabel dari dataset sebagai berikut:

1. *Intervention* – sebuah proses atau tindakan yang menjadi fokus studi klinis
  2. *Intervention Classification* – pengelompokan dari jenis *intervention* yang diterapkan
  3. *Developer and Licensing Status* – Tempat dimana penanganan berlangsung beserta dengan izin penanganannya
  4. *Trial ID* – Identitas dari pasien
  5. *Locations* – Asal negara pasien
  6. *In Monotherapy / Combination* – Jenis perawatan yang hanya diberikan 1 jenis obat saja untuk pengobatannya (*Monotherapy*) dan diberikan lebih dari 2 jenis obat secara bersamaan untuk pengobatannya (*Combination*)
  7. *Primary Completion Date* – Tanggal perawatan selesai
  8. *Intervention Type* – Jenis penanganan yang diterapkan apakah terapi menyembuhkan penyakit atau pencegahan
  9. *Trial phase* – Tahap uji klinis yang mempelajari obat atau produk biologis, berdasarkan definisi yang dikembangkan oleh Food and Drug Administration (FDA) US. Tahapan tersebut didasarkan pada tujuan studi, jumlah peserta, dan karakteristik lainnya. Ada lima fase diantaranya, Fase Awal 1 (sebelumnya terdaftar sebagai Fase 0), Fase 1, Fase 2, Fase 3, dan Fase 4. Tidak berlaku digunakan untuk menggambarkan uji coba tanpa fase yang ditentukan FDA, termasuk uji coba perangkat atau intervensi perilaku.
  10. *Registration Date* – Tanggal pendaftaran penanganan
  11. *Trial Registry – Platform* pendaftaran penanganan
  12. *Results* – Hasil dari penanganan
  13. *Status* – Status dari perekrutan penanganan
  14. *URL hyperlink* – Link detail dari penanganan setiap pasien
- Setelah mengumpulkan *dataset*, maka akan dilakukan proses *Data Selection* dimana terdapat variabel yang tidak diperlukan pada penelitian ini, sehingga

akan di-*filter* kembali sesuai dengan variabel yang dibutuhkan nantinya. Karena tujuan penelitian ini ialah untuk memprediksi apakah dengan suatu jenis penanganan bisa menyembuhkan pasien yang mengidap virus COVID-19 atau tidak, maka dari itu ada beberapa variabel yang kurang diperlukan pada penelitian ini yang bersifat sangat mendetil.

Adapun beberapa variabel yang tidak terpakai sebagai berikut:

1. *Developer and Licensing Status* – variabel ini berisi organisasi yang melakukan penanganan ini dan surat izin tidak diperlukan karena pada penelitian ini hanya membahas mengenai jenis penanganannya saja. Jikalau jenis suatu penanganan berhasil pada pasien di suatu tempat, maka penanganan itu juga akan berhasil di tempat asal manapun.

2. *Trial ID* – variabel ini berisi identitas dari pasien yang tidak diperlukan pada penelitian ini.

3. *Locations* – variabel ini berisi kewarganegaraan pasien yang bukan bagian dari penelitian ini. Jikalau jenis suatu penelitian berhasil pada pasien, maka penanganan itu juga akan berhasil di negara manapun.

4. *Primary Completion Date* – variabel ini berisi kapan pasien selesai ditangani, dimana pada penelitian ini tidak membahas mengenai berapa lama penanganan akan diberikan kepada pasien, melainkan hanya prediksi kesembuhannya saja.

5. *Trial phase* – variabel ini berisi fase ke- berapa penanganan ini dilakukan dan tidak diperlukan karena pada penelitian ini tidak membahas mengenai fase penanganan dilakukan.

6. *Registration Date* – variabel ini berisi kapan pendaftaran pasien untuk penanganannya dan juga tidak diperlukan karena pada penelitian ini tidak membahas durasi penanganannya.

7. *Trial Registry* – variabel ini berisi detil dari penanganan pasien ditangani oleh organisasi apa dan tidak diperlukan

karena pada penelitian ini tidak membahas mengenai asal organisasi penanganannya.

8. *Status* – variabel ini berisi status perekrutan pasien sudah di tahap mana dan tidak diperlukan karena pada penelitian ini semua dataset yang dipakai sudah ter-*filter* menggunakan status “*Completed*” sehingga tidak perlu dicantumkan kembali lagi.

9. *URL hyperlink* – variabel ini berisi situs yang menunjukkan detil dari setiap penanganan dari pasien yang tidak diperlukan karena pada penelitian ini hanya membahas mengenai prediksi kesembuhannya saja.

Variabel yang dipakai untuk penelitian ini sebagai berikut:

1. *Intervention* – variabel ini berisi deskripsi secara detil pengobatan apa yang digunakan dan diperlukan karena pada penelitian ini membahas kesembuhan pasien yang didasarkan dengan jenis obat-obatan yang dipakai.

2. *Intervention Classification* – variabel ini berisi jenis pengelompokan penanganannya yang lebih terstruktur datanya dibandingkan dengan hanya *Intervention* yang sangat detil. Sehingga cocok digunakan variabel ini untuk penelitian ini.

3. *In Monotherapy / Combination* – variabel ini berisi apakah terapi yang diberikan hanya satu jenis atau lebih dari satu jenis (kombinasi). Sehingga pada perhitungan penelitian ini bisa digunakan untuk mempengaruhi jumlah terapi yang diberikan ke pasien.

4. *Intervention Type* – variabel ini berisi jenis pengobatan yang diberikan itu apakah terapi dalam mengobati penyakit atau pencegahan terinfeksi penyakit. Sehingga mempengaruhi perhitungan prediksi kesembuhan pada penelitian ini.

5. *Results* – variabel ini berisi apakah terapi yang diberikan kepada pasien sukses atau tidaknya. Sehingga ini

menjadi output dari prediksi kesembuhan pada penelitian ini.

### PERSIAPAN DATA

Berikut merupakan proses dalam menguji kelayakan tersebut:

#### 1. *Data Selection* (Pemilihan Data)

Pemilihan Data merupakan proses pemilihan data-data yang berguna untuk digunakan nantinya terdapat beberapa variable-variabel yang tidak diperlukan dalam perhitungan nantinya seperti *Trial ID*, *Registration Date*, *Primary Completion Date*, *Status*, *Trial Phase*, dan *Locations*.

#### 2. *Data Cleaning* (Pembersihan Data)

Pembersihan Data merupakan proses penghapusan kesalahan data, nilai yang hilang dan data yang tidak konsisten, pada penelitian ini telah dilakukan pembersihan data yang isinya tidak relevan yaitu pada isi variabel *Intervention Classification* yang berisi nilai "Others".

Setelah melakukan beberapa proses diatas, total jumlah dataset yang diperoleh menjadi sebanyak 286 buah data dari sebelumnya sebanyak 367 buah data.

### PENGUJIAN DATA

Berikut merupakan rumus dari metode Naïve Bayes[3]:

$$P(Q|X) = \frac{P(X|Q) \cdot P(Q)}{P(X)}$$

Keterangan:

*X* : Data dengan kelas yang tidak diketahui

*Q* : Hipotesis *X* adalah kelas tertentu

$P(Q|X)$  : Probabilitas hipotesis *Q* mengacu pada *X* (*posteriori probability*)

$P(Q)$  : Probabilitas hipotesis *Q* (*prior probability*)

$P(X|Q)$  : Probabilitas *X* dalam hipotesis *Q*

$P(X)$  : Probabilitas *X*

Adapun langkah perhitungan probabilitas kesembuhan pasien yang mengidap virus COVID-19 yaitu:

1. Tahap 1:

Memisahkan 2 kelas menjadi C1 untuk "Results = Yes" dan C2 untuk "Results = No"

2. Tahap 2:

Mencari *Prior Probability* ( $P(Q)$ ):

a. Probabilitas Results dengan hasil Yes

b. Probabilitas Results dengan hasil No

3. Tahap 3:

Mencari *Likelihood* ( $Q_1 \dots Q_n$ ) dari setiap atribut

4. Tahap 4:

Mencari *Likelihood Probability* ( $P(X|Q)$ ):

a. Probabilitas Terapi ( $P(X)$ ) dengan hasil Results = Yes berdasarkan hasil penanganan pasien

b. Probabilitas Terapi ( $P(X)$ ) dengan hasil Results = No berdasarkan hasil penanganan pasien

5. Tahap 5:

Mencari *Posteriori Probability* ( $P(X|Q)$ ):

a. Mengalikan Probabilitas Terapi ( $P(X)$ ) dengan hasil Results = Yes berdasarkan hasil penanganan pasien dari *Likelihood Probability* ( $P(X|Q)$ ) dengan Probabilitas Results = Yes dari *Prior Probability* ( $P(Q)$ )

b. Mengalikan Probabilitas Terapi ( $P(X)$ ) dengan hasil Results = No berdasarkan hasil penanganan pasien dari *Likelihood Probability* ( $P(X|Q)$ ) dengan Probabilitas Results = No dari *Prior Probability* ( $P(Q)$ )

6. Tahap 6:

Membandingkan hasil dari *Posteriori Probability* ( $P(Q|X)$ ) antara Results = Yes dan No. Nilai yang lebih besar merupakan hasil dari penanganan terapi yang diberikan kepada pasien apakah sukses atau tidak dengan menggunakan metode Naïve Bayes.

### HASIL PENELITIAN

Pada bagian ini akan dibahas mengenai hasil yang diperoleh dari perhitungan dengan menggunakan pembagian data sebesar 70% untuk data training 200 buah dan 30% untuk data testing 86 buah.

Dari 70% data training sebanyak 200 buah data dan 30% data testing yaitu sebanyak 86 buah data, menghasilkan sebagian jumlah hasil prediksinya hampir menyerupai hasil aktualnya. Sehingga, kombinasi dari metode Naïve Bayes dengan data yang diperoleh tingkat akurasi adalah sebesar 96,51%.

### **Precision (Presisi)**

Untuk perhitungan *Precision* (Presisi), didapatkan dari perbandingan jumlah hasil prediksi Sukses/Gagal (*Yes/No*) yang sama dengan hasil aktualnya, dibagi dengan total prediksi Sukses/Gagal (*Yes/No*).

Pada kasus Presisi Sukses (*Yes*), didapatkan 16 hasil prediksi Sukses (*Yes*) yang hasilnya sama dengan data aktualnya, dibagi dengan total hasil prediksi Sukses (*Yes*) sebanyak 16 data, maka akan memperoleh hasil sebagai berikut:

$$= \frac{16(\text{Prediksi Sukses (Yes) yang hasilnya sama dengan data aktual})}{16(\text{total prediksi Sukses (Yes)})} \times 100\% \\ = 100\%$$

Pada kasus Presisi Gagal (*No*), didapatkan 67 hasil prediksi Gagal (*No*) yang hasilnya sama dengan data aktualnya, dibagi dengan total hasil prediksi Gagal (*No*) sebanyak 70 data, maka akan memperoleh hasil sebagai berikut:

$$= \frac{67(\text{Prediksi Gagal (No) yang hasilnya sama dengan data aktual})}{70(\text{total prediksi Gagal (No)})} \times 100\% \\ = 95,71\%$$

### **Recall (Sensitifitas)**

Untuk perhitungan *Recall* (Sensitifitas), didapatkan dari perbandingan jumlah hasil prediksi Sukses/Gagal (*Yes/No*) yang sama dengan hasil aktualnya, dibagi dengan total hasil Sukses/Gagal

(*Yes/No*) yang aktual dari data *testing* yang digunakan.

Pada kasus Sensitifitas Sukses (*Yes*), didapatkan 16 hasil prediksi Sukses (*Yes*) yang hasilnya sama dengan data aktualnya, dibagi dengan total hasil Sukses (*Yes*) yang aktual dari data testing yang digunakan sebesar 19 data, maka akan memperoleh hasil sebagai berikut:

$$= \frac{16(\text{Prediksi Sukses (Yes) yang hasilnya sama dengan data aktual})}{19(\text{total hasil Sukses (Yes) aktual di data testing})} \times 100\% \\ = 84,21\%$$

Pada kasus Sensitifitas Gagal (*No*), didapatkan 67 hasil prediksi Gagal (*No*) yang hasilnya sama dengan data aktualnya, dibagi dengan total hasil Gagal (*No*) yang aktual dari data testing yang digunakan sebesar 67 data, maka akan memperoleh hasil sebagai berikut:

$$= \frac{67(\text{Prediksi Gagal (No) yang hasilnya sama dengan data aktual})}{67(\text{total hasil Gagal (No) aktual di data testing})} \times 100\% \\ = 100\%$$

### **Accuracy (Akurasi)**

Untuk perhitungan *Accuracy* (Akurasi), didapatkan dari jumlah hasil prediksi Sukses (*Yes*) yang sama dengan hasil aktualnya, dibagi dengan total data testing yang digunakan. Kemudian dijumlahkan dengan jumlah hasil prediksi Gagal (*No*) yang sama dengan hasil aktualnya, dibagi dengan total data testing yang digunakan.

Kemudian didapatkan 16 hasil prediksi Sukses (*Yes*) yang hasilnya sama dengan data aktualnya, dibagi dengan total data testing yang digunakan sebanyak 86 data, maka akan memperoleh hasil sebagai berikut:

$$= \frac{16(\text{Prediksi Sukses (Yes) yang hasilnya sama dengan data aktual})}{86(\text{total data testing})} \times 100\% \\ = 18,60\%$$

Dalam penelitian ini didapatkan 67 hasil prediksi Gagal (*No*) yang hasilnya sama dengan data aktualnya, dibagi dengan total data testing yang digunakan sebanyak 86 data, maka akan memperoleh hasil sebagai berikut:

$$= \frac{67(\text{Prediksi Gagal (No) yang hasilnya sama dengan data aktual})}{86(\text{total data testing})} \times 100\% = 77,91\%$$

Maka jika dijumlahkan dari masing-masing perbandingan yang telah diperhitungkan akan mendapatkan tingkat akurasi sebesar  $= 18,60\% + 77,91\% = 96,51\%$ .

Jenis terapi yang diberikan kepada pasien dalam *dataset* ini sangat variatif, sehingga pelatihan untuk data training akan sedikit lebih menantang karena luasnya perbedaan isi pada variabel Intervention, tetapi dengan adanya perbedaan yang luas maka semakin banyak alternatif dalam memprediksi tingkat kesembuhan pasiennya dengan jenis terapi yang berbeda-beda.

Berikut merupakan jenis terapi yang sukses menyembuhkan pasien secara data aktual yaitu jenis terapi *Acalabrutinib (Calquence)*, *Ampion*, *Anakinra (Kineret)*, *anti-SARS-CoV-2 IgY*, *ArtemiC*, *Asbrip*, *Aspirin*, *Astegolimab (RG-6149)*, *AT-001*, *Atovaquone*, *AV-COVID-19*, *Aviptadil (Zyesami)*, *Azithromycin*, *Bamlanivimab (LY3819253, LY-CoV555)*, *Bardoxolone Methyl (RTA 402)*, *Baricitinib (Olumiant)*, *Berberine*, *Bevacizumab (Avastin)*, *BGB-DXP593*, *Brensocatic (INS1007)*, *Brequinar*, *BRII-198*, *Bromhexine hydrochloride*, *BTL-TML001*, *Camostat mesylate*, *Canakinumab (Ilaris)*, *CAPEo (Cretan IAMA)*, *CD24Fc*, *CERC-002*, *Cetylpyridinium chloride*, *Chloroquine*, *Cholecalciferol (Vitamin D3)*, *Clarithromycin*, *Clindamycin*, *Colchicine*, dan *Convalescent plasma*.

Berikut merupakan hasil prediksi jenis terapi menyembuhkan pasien yang sukses menggunakan metode Naïve Bayes dengan 30% data testing sebanyak 86 data yang dipilih dari 286 data secara acak dan hasilnya sama dengan data aktualnya yaitu jenis terapi *Azithromycin*,

*Doxycycline*, *Acalabrutinib (Calquence)*, *Astegolimab (RG-6149)*, *Atovaquone*, *Bardoxolone Methyl (RTA 402)*, *Camostat mesylate*, *Canakinumab (Ilaris)*, *Chloroquine*, *Colchicine*, *Crizanlizumab (Adakveo)*, dan *Human Immunoglobulin*. Sehingga jenis terapi yang dihasilkan cukup variatif dan memiliki banyak alternatif dalam mengatasi penyakit COVID-19 jika digabungkan dengan data aktualnya.

Maka dari itu, diperoleh kesimpulan perhitungan menggunakan metode Naïve Bayes pada penelitian ini akan menghasilkan tingkat keakuratan kesembuhan COVID-19 sebesar 96,51% yang artinya hasil perhitungan prediksi sukses dan gagal dalam suatu terapi yang diberikan kepada pasien dapat dipertanggungjawabkan sebagai acuan data dalam penelitian berikutnya yang lebih mendetail. Tingkat keakuratan ini masih dapat ditingkatkan dengan cara memperoleh jumlah data yang lebih banyak lagi sehingga dapat meningkatkan akurasi ketika melakukan pengujian data.

## KESIMPULAN

Dari hasil penelitian dan perhitungan yang dilakukan menggunakan metode Naïve Bayes, dapat ditarik beberapa kesimpulan seperti:

1. Hasil perhitungan *precision* (presisi) yang didapatkan dari perbandingan jumlah hasil prediksi Sukses/Gagal (Yes/No) yang sama dengan hasil aktualnya, dibagi dengan total prediksi Sukses/Gagal (Yes/No) menghasilkan nilai sebesar 100% untuk Presisi Sukses (Yes) dan 95,71% untuk Presisi Gagal (No).
2. Perhitungan *recall* (sensitifitas) yang didapatkan dari perbandingan jumlah hasil prediksi Sukses/Gagal (Yes/No) yang sama dengan hasil aktualnya, dibagi dengan total hasil Sukses/Gagal (Yes/No) yang aktual dari *data testing* yang

digunakan yang menghasilkan nilai sebesar 84,21% untuk Sensitifitas Sukses (Yes) dan 100% untuk Sensitifitas Gagal (No).

3. Perhitungan *Accuracy* (Akurasi) yang didapatkan dari jumlah hasil prediksi Sukses (*Yes*) yang sama dengan hasil aktualnya, dibagi dengan total *data testing* yang digunakan. Kemudian dijumlahkan dengan jumlah hasil prediksi Gagal (*No*) yang sama dengan hasil aktualnya, dibagi dengan total *data testing* yang digunakan. Maka akan menghasilkan tingkat keakuratan sebesar 96,51%.

Adapun saran untuk penelitian berikutnya adalah melibatkan faktor-faktor lain diluar faktor terapi seperti faktor lingkungan, *antibody*, nutrisi, kegiatan sehari-hari, obat yang dikonsumsi, frekuensi olahraga, dan sebagainya.

#### DAFTAR PUSTAKA

- [1] Matta, D. M., & Saraf, M. K. (2020). Prediction of COVID-19 using Machine Learning Techniques. Karlskrona: Blekinge Institute of Technology.
- [2] Adrian Schmid, A. W. (2019). Monotherapy versus combination therapy for multidrug-resistant Gram-negative infections: Systematic Review and Meta-Analysis. *Scientific Reports Nature Research*, 2.
- [3] Aji Prasetya Wibawa, A. C. (2019). Naïve Bayes Classifier for Journal Quartile Classification. *International Journal of Recent Contributions from Engineering Science & IT (iJES)*, 94-96.
- [4] Excellence, N. I. (2020). National Institute for Health and Care Excellence. Retrieved from National Institute for Health and Care Excellence: <https://www.nice.org.uk/covid-19/rapid-c19>
- [5] Gardner, L. (2020, January 23). Coronavirus COVID-19 (2019-nCov). Retrieved from COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU): <https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>