

PREDIKSI KASUS COVID-19 INDONESIA MENGGUNAKAN METODE NAÏVE BAYES

PREDICTING INDONESIAN'S COVID-19 CASES USING NAÏVE BAYES METHOD

Info artikel Diterima: 21 Agustus 2022 Direvisi: 30 November 2022 Disetujui: 28 Desember 2022

Dafid^{1*}, Veratiwi²

¹ Universitas Multi Data Palembang, Sumatera Selatan, Indonesia

² Poltekkes Kemenkes Palembang, Sumatera Selatan, Indonesia

(*email penulis korespondens: dafid@mdp.ac.id)

ABSTRAK

Latar Belakang: COVID-19 merupakan virus yang menyerang sistem pernapasan. Dampak dari COVID-19 sangat signifikan bagi kehidupan masyarakat Indonesia yang terjadi hampir di semua bidang. Berdasarkan data dari Kementerian Kesehatan (Kemenkes) melalui laman <https://www.kemkes.go.id/> per 18 November 2022, kasus penyebaran dan kasus terinfeksi oleh virus ini terus bertambah. Hal ini menjadi tantangan besar dalam menghadapi COVID-19 terutama bagi pembuat kebijakan yang harus mengambil keputusan dari berbagai macam informasi yang masih belum sempurna. Penelitian ini bertujuan untuk memprediksi kasus COVID-19 di Indonesia sehingga para pembuat kebijakan dapat mengambil keputusan yang sesuai baik itu untuk pencegahan maupun untuk penyusunan strategi dalam menangani kasus COVID-19.

Metode: Dalam penelitian ini bahan yang digunakan untuk kemudian diolah dan menjadi acuan adalah dataset yang dikompilasi secara *time series* dari [covid19.go.id](https://www.covid19.go.id/), [kemendagri.go.id](https://www.kemendagri.go.id/), [bps.go.id](https://www.bps.go.id/), dan [bnpp-inacovid19.hub.arcgis.com](https://www.bnpp-inacovid19.hub.arcgis.com/) berupa data kasus covid-19 di Indonesia yang *diupdate* dalam 2 bulan terakhir. Metode yang digunakan untuk melakukan prediksi menggunakan dataset yang telah disiapkan adalah *Naïve Bayes*.

Hasil: Prediksi yang dilakukan dengan menggunakan metode *Naïve Bayes* memberikan hasil prediksi dengan mengklasifikasi kasus baru yang terjadi berdasarkan variabel independen yang telah ditentukan. Nilai akurasi yang dihasilkan yaitu sebesar 89%.

Kesimpulan: Metode *Naïve Bayes* mampu memberikan hasil prediksi dengan tepat dengan tingkat akurasi yang dihasilkan sangat tinggi. Hal ini dipengaruhi oleh faktor banyaknya data yang disediakan dan juga banyaknya variabel yang digunakan untuk melakukan prediksi.

Kata kunci : COVID-19, *Naïve Bayes*, Akurasi

ABSTRACT

Background: COVID-19 is a one of virus that attacks the respiratory system. The impact of COVID-19 is very significant for the life of the Indonesian people which occurs in almost all fields. Based on data from the Ministry of Health (Kemenkes) via the website <https://www.kemkes.go.id/> as of November 20 2022, cases of spread and cases of infection with this virus continue to increase. This is a big challenge in dealing with COVID-19, especially for policy makers who have to make decisions from various kinds of information that is still not perfect. This study aims to predict cases of COVID-19 in Indonesia so that policy makers can make appropriate decisions both for prevention and for developing strategies in dealing with cases of COVID-19.

Methods: In this study, the materials used to be processed and used as a reference are datasets compiled in a time series from [covid19.go.id](https://www.covid19.go.id/), [kemendagri.go.id](https://www.kemendagri.go.id/), [bps.go.id](https://www.bps.go.id/), and [bnpp-inacovid19.hub.arcgis.com](https://www.bnpp-inacovid19.hub.arcgis.com/). in the form of data on covid-19 cases in Indonesia which were updated in the last 2 months. The method used to make predictions using the prepared dataset is *Naïve Bayes*.

Results: Predictions made using the *Naïve Bayes* method provide prediction results by classifying new cases based on predetermined independent variables. The resulting accuracy value is 89%

Conclusion: The *Naïve Bayes* method can provide precise prediction results with a very high level of accuracy. This is influenced by the amount of data provided and the number of variables used to make predictions.

Keywords : COVID-19, *Naïve Bayes*, Accuracy

PENDAHULUAN

COVID-19 telah menjadi salah satu isu kesehatan nasional bahkan secara global. COVID-19 merupakan virus yang menyerang sistem pernapasan manusia dimana pada kasus yang parah, covid-19 dapat menyebabkan *pneumonia*, sindrom pernapasan akut bahkan sampai dengan kematian. COVID-19 telah melanda hampir diseluruh dunia dengan data dari WHO per tanggal 21 April 2020 total kasus yang terinfeksi virus ini mencapai 2.397.217 dengan kasus kematian mencapai 162 kasus dari seluruh dunia. COVID-19 terdeteksi pertama kali di Indonesia pada tanggal 2 Maret 2020 yang berjumlah 2 kasus (1,2). Sejak pertama kali kemunculannya hingga saat ini angka penyebaran kasus COVID-19 mengalami fluktuasi. Data 31 Maret 2020 menunjukkan kasus yang terkonfirmasi berjumlah 1.528 kasus dan 136 kasus kematian (3). Tingkat mortalitas COVID-19 di Indonesia sebesar 8,9% dimana angka ini merupakan yang tertinggi di Asia Tenggara (4,5). Tingginya angka penyebaran COVID-19 memberikan dampak sosial dan ekonomi di Indonesia. Untuk dampak ekonomi yang terjadi salah satunya adalah menurunnya pertumbuhan ekonomi dan meningkatnya angka pengangguran dan kemiskinan (1).

Untuk mengatasi hal ini pemerintah telah mengambil langkah-langkah untuk menekan peningkatan kasus COVID-19 yaitu dengan memberlakukan pembatasan kegiatan masyarakat dan pemberian vaksin. Dalam hal ini pemerintah membutuhkan informasi lebih awal dalam membuat keputusan maupun dalam

METODE

Penelitian ini merupakan penelitian prediktif yang memprediksikan atau memperkirakan apa yang akan terjadi atau berlangsung pada saat yang akan datang berdasarkan hasil analisis keadaan saat ini. Analisis data yang digunakan adalah analisis Prediktif Analitik yang mempelajari mengenai hubungan-hubungan antar variabel dan kemudian membuat suatu model statistik yang dapat memprediksi suatu nilai dari kejadian baru dan kejadian masa depan. Penelitian ini dimulai dengan melakukan studi literatur tentang perkembangan kasus COVID-19 di Indonesia

penyusunan strategi. Beberapa penelitian sebelumnya yang telah dilakukan oleh para peneliti hanya dilakukan secara lokal dengan menggunakan beberapa algoritma klasifikasi berbeda dan memberikan hasil yang perlu disempurnakan. Metode yang digunakan adalah *K-Nearest Neighbors*, *Backpropagation* dan *Fuzzy Tsukamoto*, *Exponential Smothing*, *Naive Bayes*, *Random Forest*, *Monte Carlo*, *Arima*, Regresi Linear Berganda. Hasil penelitian hanya menunjukkan hasil perhitungan dengan menggunakan metode prediksi tertentu tanpa menunjukkan akurasi dari model prediksi yang dibuat (6–10). Namun ada beberapa penelitian yang telah dilakukan secara nasional dengan hasil yang masih dan perlu ditingkatkan dan perlu diuji tingkat akurasinya (1,11–20). Selain itu data yang digunakan dalam penelitian tersebut perlu ditelaah lebih lanjut mengingat hasil penelitian sangat tergantung dengan data yang digunakan. Data yang digunakan tidak begitu representatif untuk memprediksi kasus COVID-19 di Indonesia dalam hal jumlah variabel yang digunakan dan banyaknya data. Oleh karena itu perlu dilakukan penelitian lebih lanjut yang dapat memberikan solusi untuk mencegah kenaikan COVID-19 di Indoensia.

Penelitian ini dilakukan dengan tujuan untuk memprediksi kasus COVID-19 di Indonesia yang akan memberikan informasi lebih awal terkait dengan jumlah kasus COVID-19 di Indonesia secara akurat. Dengan adanya informasi lebih awal tersebut para pembuat kebijakan dapat mengambil keputusan dengan lebih mudah dan lebih cepat dan tepat sehingga dengan demikian dampak sosial ataupun ekonomi dapat diminimalisir.

dan metode yang digunakan untuk membuat model prediksinya. Studi literatur dilakukan dari jurnal serta web terkait. Selanjutnya melakukan pengumpulan data yang berkaitan dengan kasus COVID-19. Data yang digunakan merupakan data sekunder yang merupakan data perkembangan COVID-19 di Indonesia mulai dari bulan Maret 2020. Data ini diperbarui terakhir bulan September 2022 yang merupakan kompilasi data dari berbagai sumber yaitu covid19.go.id, kemendagri.go.id, bps.go.id, dan bnpb-inacovid19.hub.arcgis.com yang berupa data *time series*. Dengan semakin banyaknya data maka akurasi prediksi yang dihasilkan semakin tinggi. Untuk memperoleh

data tersebut dapat diunduh melalui repository *Kaggle* dengan nama *COVID-19 Indonesia Dataset* yang dapat diakses melalui alamat web <https://www.kaggle.com/hendratno/covid19->

indonesia yang berupa file .csv yang terdiri dari 31.823 baris dan 38 variabel dengan nilai numerik dan nominal yang deskripsinya dapat dilihat pada Tabel 1.

Tabel 1. Deskripsi COVID-19 Indonesia Dataset

Variabel	Deskripsi
Date	Tangga pelaporan data
Location ISO Code	Kode lokasi berdasarkan standar ISO
Location	Nama lokasi
New Cases	Kasus positif harian baru
New Deaths	Kasus harian kematian baru
New Recovered	Jumlah sembuh harian baru
New Active Cases	Kasus aktif harian baru
Total Cases	Total kasus positif sampai tanggal terkait
Total Deaths	Total kematian sampai tanggal terkait
Total Recovered	Total sembuh sampai tanggal terkait
Total Active Cases	Total kasus aktif sampai tanggal terkait
Location Level	Level lokasi
City or Regency	Kota atau Kabupaten
Province	Nama provinsi dari lokasi
Country	Nama negara dari lokasi
Continent	Nama benua dari lokasi
Island	Nama pulau dari lokasi
Time Zone	Zona waktu dari lokasi
Special Status	Status khusus dari lokasi
Total Regencies	Total kabupaten di lokasi
Total Cities	Total kota di lokasi
Total Districts	Total kecamatan di lokasi
Total Urban Villages	Total kelurahan di lokasi
Total Rural Villages	Total desa di lokasi
Area (km²)	Area lokasi (km ²)
Population	Populasi di lokasi
Population Density	Kepadatan populasi di lokasi
Longitude	Longitude dari lokasi
Latitude	Latitude dari lokasi
New Cases per Million	Jumlah kasus baru per 1 juta
Total Cases per Million	Total kasus per 1 juta
New Deaths per Million	Jumlah kematian baru per 1 juta
Total Deaths per Million	Total kematian per 1 juta
Total Deaths per 100rb	Total kematian per 100 ribu
Case Fatality Rate	Tingkat kematian kasus
Case Recovered Rate	Tingkat kesembuhan kasus
Growth Factor of New Cases	Faktor pertumbuhan kasus baru
Growth Factor of New Deaths	Faktor pertumbuhan kematian baru

Data yang sudah diperoleh selanjutnya diolah dan dianalisis dengan menggunakan metode *Naïve Bayes*. Metode ini dipilih karena banyak digunakan para peneliti pada bidang kesehatan yang menghasilkan tingkat akurasi yang tinggi (21). Metode *Naïve Bayes* merupakan metode prediksi probabilitas dari suatu data yang disajikan dengan mencari peluang terbesar dari kemungkinan dalam klasifikasi berdasarkan frekuensi tiap klasifikasi terhadap data *training*.

HASIL

Penelitian ini menggunakan tool berupa perangkat lunak Python yang *powerful* untuk mengolah dan menganalisa data serta menghasilkan suatu model prediksi. Data yang

Adapun rumus perhitungan dari *Naïve Bayes* yaitu: $P(X|Y) = P(Y|X)P(X)/P(Y)$ dengan
Y= data dengan kelas yang belum diketahui
X= hipotesis data Y merupakan suatu kelas spesifik
 $P(X|Y)$ = probabilitas hipotesis X berdasarkan kondisi Y
 $P(Y|X)$ = probabilitas Y berdasarkan kondisi pada hipotesis X
 $P(X)$ = probabilitas hipotesis X
 $P(Y)$ = probabilitas Y.

telah diunduh selanjutnya akan diimpor ke Python untuk dilakukan proses pembacaan data dan selanjutnya akan melalui fase analisis data, perancangan data, pelatihan dan pengujian data. Hasil pembacaan data dapat dilihat pada Gambar 1 berikut ini:

```
[31817 rows x 38 columns]
['Date' 'Location ISO Code' 'Location' 'New Cases' 'New Deaths'
 'New Recovered' 'New Active Cases' 'Total Cases' 'Total Deaths'
 'Total Recovered' 'Total Active Cases' 'Location Level' 'City or Regency'
 'Province' 'Country' 'Continent' 'Island' 'Time Zone' 'Special Status'
 'Total Regencies' 'Total Cities' 'Total Districts' 'Total Urban Villages'
 'Total Rural Villages' 'Area (km2)' 'Population' 'Population Density'
 'Longitude' 'Latitude' 'New Cases per Million' 'Total Cases per Million'
 'New Deaths per Million' 'Total Deaths per Million'
 'Total Deaths per 100rb' 'Case Fatality Rate' 'Case Recovered Rate'
 'Growth Factor of New Cases' 'Growth Factor of New Deaths']
```

Gambar 1. Import data COVID-19 Indonesia Dataset ke Python

Pada fase analisis, ditentukan variabel yang akan menjadi target prediksi sebagai variabel dependen yaitu *New Case* yang merupakan variabel yang berisi kenaikan covid-19 perhari dari seluruh provinsi di Indonesia. Lalu dari sisa variabel yang ada akan menjadi variabel independen. Selanjutnya dataset ini dibagi menjadi data *training* dan data *testing*. Data *training* adalah data yang digunakan sebagai dasar untuk membuat model prediksi. Sedangkan data *testing* merupakan data yang digunakan untuk mengukur

ketepatan prediksi dari model yang dibangun. Dalam hal ini pembagiannya adalah 90% data *training* dan 10% adalah data *testing* yang berarti dari 31.823 data, 28.641 adalah data *training* dan 3.182 adalah data *testing*. Adapun tipe sampling yang digunakan adalah *stratified sampling*. Ada 4 variabel yang digunakan dalam Python untuk menentukan proses pembagian data dengan menggunakan fungsi *train_test_split* yaitu *x_train*, *x_test*, *y_train*, *y_test* dengan potongan kode yang ditunjukkan pada Gambar 2 berikut ini.

```
# Import train_test_split function
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.1, random_state = 0)
```

Gambar 2. Pembagian dataset menjadi data *training* dan data *testing* dalam Python

Setelah menentukan variabel independen dan variabel dependen dan telah dilakukan pembagian data selanjutnya akan dilakukan analisis menggunakan metode *Naive Bayes* untuk melihat hasil prediksi. Tahap berikutnya dilakukan pembuatan model *Naive-Bayes* menggunakan data *training*. Dari model yang sudah dibuat dapat dilakukan prediksi dengan menggunakan data testing yang kita peroleh sebelumnya.

```
#Import model Gaussian Naive Bayes
from sklearn.naive_bayes import GaussianNB
#Membuat model Naive bayes
modelnb = GaussianNB()
#Memasukkan data training pada fungsi klasifikasi naive bayes
nbtrain = modelnb.fit(x_train, y_train)
#Memprediksi hasil data testing
y_pred = nbtrain.predict(x_test)
#Menghitung tingkat akurasi
print(accuracy_score(y_test, y_pred))
```

Gambar 3. Perhitungan hasil prediksi dan tingkat akurasi dalam Python

PEMBAHASAN

Berdasarkan dari hasil penelitian yang telah dilakukan menunjukkan hasil prediksi yang diperoleh menunjukkan tingkat akurasi yang tinggi. Bila dibandingkan dengan hasil penelitian yang dilakukan oleh peneliti sebelumnya, hasil penelitian ini lebih tinggi kurang lebih ada peningkatan 10% dimana hasil sebelumnya hanya berkisar antara 65%-75%. Hal ini dapat dilihat dari hasil penelitian yang dilakukan oleh Suci yaitu sebesar 65% (12). Sedangkan peneliti Rayuwati melaporkan hasil penelitiannya sebesar 60% (7). Berbeda dengan yang dilakukan oleh Andi dan Rizka dengan hasil penelitian berturut-turut sebesar 70% dan 75% (10,19). Hasil ini juga melengkapi dari penelitian sebelumnya yang tidak menunjukkan tingkat akurasi dari model prediksi yang dibuat. Hal ini juga diperkuat dengan banyaknya data yang digunakan dengan peningkatan jumlah data sebesar 200% lebih dibandingkan dengan penelitian sebelumnya dimana banyaknya data

DAFTAR PUSTAKA

1. Joyosemito IS. Pendekatan Pemodelan Matematika Dinamis Dalam Analisis Prediksi COVID-19 Sebagai Masukan Kebijakan Pemerintah Indonesia.

Untuk menghitung nilai keakurasian maka dibandingkan nilai prediksi dengan nilai testing dengan kode yang dapat dilihat pada Gambar 3. Dari hasil perhitungan dengan menggunakan tools aplikasi *Python* dengan kode `print(accuracy_score(y_test, y_pred))` seperti yang ditunjukkan pada Gambar 3 diperoleh nilai prediksi sebesar **0.8927168198138277** atau dengan kata lain sebesar 89%.

sebelumnya berkisar antara 200-500 data sehingga data yang digunakan dalam penelitian ini lebih representatif karena hasil kompilasi dari beberapa sumber data COVID-19 yang ada di Indonesia. Begitu juga dengan banyaknya variabel yang digunakan turut andil dalam menentukan akurasi dari prediksi yang dihasilkan.

KESIMPULAN DAN SARAN

Dari pembuatan model prediksi dan perhitungan yang dilakukan dapat disimpulkan bahwa metode *Naive Bayes* menghasilkan prediksi dengan tingkat akurasi yang tinggi sesuai dengan hasil penelitian dari peneliti sebelumnya. Faktor yang mendukung hasil penelitian diantaranya adalah banyaknya data dan banyaknya variabel yang digunakan.

Untuk meningkatkan akurasi dari model *Naive-Bayes* yang digunakan dapat diterapkan proses lain seperti *preprocessing*, seleksi variabel maupun metode data mining lainnya.

2021;4(1):1–16.

2. WHO. Coronavirus disease 2019 (COVID-19) Situation Report– 42 [Internet]. [cited 2020 Sep 10]. Available from: <https://www.who.int/docs/default->

- source/coronaviruse/situation-reports/20200302-sitrep-42-covid-19.pdf?sfvrsn=224c1add_2
3. Kementerian Kesehatan Republik Indonesia. Info Infeksi Emerging Kementerian Kesehatan RI [Internet]. [cited 2022 Sep 10]. Available from: <https://infeksiemerging.kemkes.go.id/>
 4. WHO. Coronavirus disease 2019 (COVID-19) Situation Report–54 [Internet]. [cited 2022 Sep 15]. Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200314-sitrep-54-covid-19.pdf?sfvrsn=dcd46351_2
 5. WHO. Coronavirus disease 2019 (COVID-19) Situation Report–70 [Internet]. [cited 2022 Sep 10]. Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200330-sitrep-70-covid-19.pdf?sfvrsn=7e0fe3f8_2
 6. Sinaga D, Ginting EF, Hutasuhut M. Implementasi Data Mining Dalam Memprediksi Penyebaran Covid-19 Di Kecamatan Namorambe Menggunakan Metode Regresi Linear Berganda. 2020;3(9).
 7. Rayuwati, Husna Gemasih, Irma Nizar. Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid. Jural Ris Rumpun Ilmu Tek. 2022;1(1):38–46.
 8. Ainiyah L, Bansori M. Prediksi Jumlah Kasus COVID-19 Menggunakan Metode Autoregressive Integrated Moving Average. J Sains Dasar. 2021;10(2):62–8.
 9. Purnamasari Y, Kunang YN. Pemanfaatan Data Mining Dalam Memprediksi Kasus Positif Covid-19 Di Kota Palembang Menggunakan. 2021;1(2):118–28.
 10. Risal AAN, Yusuf NI, Kaswar AB, Adiba F. Penerapan Data Mining dalam Mengklasifikasikan Tingkat Kasus Covid-19 di Sulawesi Selatan Menggunakan Algoritma Naive Bayes. Indones J Fundam Sci. 2021;7(1):18–28.
 11. Dian Arianto FS, Noviyanti. Prediksi Kasus Covid-19 Di Indonesia Menggunakan Metode Backpropagation Dan Fuzzy Tsukamoto. 2020;4(1):120–7.
 12. Anggraini S, Akbar M, Wijaya A, Syaputra H, Sobri M. Klasifikasi Gejala Penyakit Coronavirus Disease 19 (COVID-19) Menggunakan Machine Learning. J Softw Eng Ampera. 2021;2(1):57–68.
 13. Waslin TTA, Sulaiman OK, Haramaini T. Aplikasi Prakiraan Perkembangan Covid-19 Di Indonesia Menggunakan Metode Single Exponential Smoothing Berbasis Web. J Media Inform Budidarma. 2022;6(3):1509.
 14. Aser W, Samosir H, Gantini T. Analisis Dataset COVID-19 menggunakan Algoritma KNN dan Random Forest. J Strateg - J Maranatha [Internet]. 2022;4(1):58–69. Available from: <https://mail.strategi.it.maranatha.edu/index.php/strategi/article/view/325>
 15. Gibran CM, Setiyawati S, Liantoni F. Prediksi Penambahan Kasus Covid-19 di Indonesia Melalui Pendekatan Time Series Menggunakan Metode Exponential Smoothing. J Inform Univ Pamulang. 2021;6(1):112.
 16. Parhusip HA. Study on COVID-19 in the World and Indonesia Using Regression Model of SVM, Bayesian Ridge and Gaussian. J Ilm Sains. 2020;20(2):49.
 17. Widjaya JS, Agushinta R D, Puspita Sari SR. Sistem Prediksi Jumlah Pasien Covid-19 Menggunakan Metode Trend Least Square Berbasis Web. Sistemasi. 2021;10(1):39.
 18. Yuwono NR, Yulianto S. Perbandingan Berbagai Metode Exponential Smoothing Untuk Peramalan Covid Di Indonesia. IT-EXPLORE (Jurnal Penerapan Teknol Inf dan Komunikasi). 2022;01(02):155–65.
 19. Dahlia R, Wuryani N, Hadiani S, Gata W, Selawati A, Komputer FI, et al. Penerapan Data Mining Terhadap Data Covid - 19. J Inform. 2021;21(1):44–52.
 20. Prasetyowati SAD, Ismail M, Badieah B. Implementation of Least Mean Square Adaptive Algorithm on Covid-19 Prediction. JUITA J Inform. 2022;10(1):139.
 21. Tomar D. A survey on Data Mining approaches for Healthcare A survey on Data Mining approaches for Healthcare. 2014;(May).

