

PERBANDINGAN MODEL RUNTUN WAKTU DAN PREDIKSI JUMLAH KASUS COVID-19 DI INDONESIA

Bakti Siregar, F. Anthon Pangruruk, Simon Prananta Barus

Universitas Matana, Indonesia

Email: siregar.bakti@matanauniversity.ac.id, antpangruruk@matanauniversity.ac.id,
simon.barus@matanauniversity.ac.id

Abstrak

Berbagai lembaga survei, penelitian dan media telah mengemukakan bahwa hampir semua negara mengalami gejala peningkatan jumlah kasus COVID-19 dalam beberapa periode. Dalam hal ini, Indonesia juga termasuk menjadi salah satu negara yang mengalami gelombang naik turun yang cukup serius bahkan sebelum terjadinya kekebalan kelompok (*Herd Immunity*) seperti. Pemerintah terus berupaya untuk menekan penyebaran COVID-19 tersebut dengan menerapkan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) yang dikategorikan berdasarkan level satu hingga level empat untuk menangani permasalahan yang ada, baik dari sisi sosial, ekonomi dan kesehatan. Dengan demikian, peramalan terkait informasi mengenai puncak dari terjadinya kasus, serta ramalan kapan berakhirnya pandemic COVID-19 menjadi sangat penting bagi pemerintah. Sehubungan dengan itu, tujuan dari penelitian ini adalah menemukan model *time series* terbaik yang dapat digunakan untuk meramalkan jumlah kasus COVID-19 di Indonesia dengan cara membandingkan berbagai model runtun waktu dengan konsep *machine learning*. Hasil penelitian ini, memperlihatkan bahwa model PROPHET adalah model terbaik dari sebelas model lainnya, dibuktikan dengan nilai R-square dan MAPE. Dengan demikian, model PROPHET ini dapat digunakan untuk meramalkan kemungkinan jumlah kasus COVID-19 di Indonesia dan hasil peramalan tersebut dapat dijadikan referensi ilmiah bagi pemerintah dalam menentukan kebijakan penanganan yang lebih baik.

Kata kunci: COVID-19; *Time Series*, *Machine Learning*; Evaluasi Model; Peramalan

Abstract

There are various institutions of the survey, research, and media have found that almost all countries have experienced an increase in the number of COVID-19 cases over several periods. Indonesia is also one of the countries that experienced serious ups and downs even before the occurrence of Herd Immunity such as. The Indonesian government continuously suppering COVID-19 through the Implementation of Community Activity Restrictions (PPKM) which are categorized based on levels one to level four to deal with existing problems, in terms of social, economic, and health aspects. Therefore, forecasting related to the number of occurrences and when the COVID-19 pandemic will end is very important for the

government. In this regard, this study aims to find the best time series model that can be used to predict the number of COVID-19 cases in Indonesia by comparing various time series models with machine learning concepts. The results of this study show that the PROPHET model is the best model out of eleven other models, as evidenced by the R-square and MAPE values. Accordingly, the PROPHET model can be used to predict the possible number of COVID-19 cases in Indonesia and the forecasting results can be used as a scientific reference for the government in determining better handling policies.

Keywords: COVID-19; Time Series, Machine Learning; Model Evaluation; Forecasting.

Pendahuluan

Peramalan jumlah kasus COVID-19 merupakan salah satu tindakan sangat penting dilakukan bagi penanggung jawab atau pengambil kebijakan di suatu negara (Darmawansyah, Indar, Arifin, & Balqis, 2021). Hal tersebut perlu dilakukan karena puncak dari terjadinya kasus hingga terkait kapan akan berakhirnya COVID-19 bisa saja akan mengalami satu, dua, maupun beberapa gelombang. Beberapa penelitian telah membuktikan bahwa kasus COVID-19 mengikuti pola tertentu yang mana pola ini berdasarkan pada transmisi yang dinamis dari epidemic (Diksa, N, 2021). Penelitian yang dilakukan oleh (Pakan, 2020) telah mengemukakan bahwa setiap epidemi di suatu negara telah muncul dengan proporsi yang berbeda dari waktu ke waktu, terutama ketika perubahan periode cuaca dan penyebaran virus selama periode tersebut sehingga membentuk pola nonlinear. Senada dengan penelitian yang dilakukan oleh (Harini, 2020) yang menggunakan analisis pendekatan time series dan ditemukan bahwa model sebaran COVID-19 di Indonesia mengikuti model double exponential.

Pada ruang lingkup penelitian yang lebih luas yaitu mengenai kasus penyebaran COVID-19 di dunia dapat dilakukan dengan menggunakan model logistik regresi nonlinier (Wang, Zheng, Li, & Zhu, 2020). Tak jauh berbeda dengan peneliti sebelumnya, (Satrio, Darmawan, Nadia, & Hanafiah, 2021) menggunakan model ARIMA dan PROPHET untuk melakukan analisis deret waktu dan peramalan penyakit virus corona di Indonesia dan ditemukan bahwa model PROPHET memiliki akurasi yang baik dalam memprediksi kasus yang dikonfirmasi dengan presisi 91%, sedangkan ARIMA bahkan tidak melewati presisi 50%. Model yang tak kalah menarik untuk dipertimbangkan dalam penelitian ini adalah jaringan saraf Nonlinear Autoregressive (NAR) dirancang untuk meramalkan deret waktu dari nilai masa lalu (Sayfeddine, 2014). Jaringan berbasis NAR ini telah terbukti dapat digunakan untuk model dari data training dan memprediksi nilai masa depan (Kirbas, 2018).

Langkah yang ditempuh pemerintah Indonesia yaitu untuk menekan penyebaran COVID-19 melalui implementasi kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) yang pada akhirnya sangat berpengaruh terhadap peningkatan jumlah penduduk miskin yang secara signifikan (Mawar, Andriyani, Gultom, & Ketiara, 2021) Kondisi ini berpotensi menjadi pemicu permasalahan di rumah tangga dan mengakibatkan angka perceraian karena himpitan ekonomi. Selain itu, angka pengangguran yang semakin melonjak sangat dikawatirkan akan menyebabkan tindakan

kriminalitas juga meningkat. Dampak ekonomi dan sosial ini tentunya menjadi tantangan pemerintah dalam mengevaluasi kebijakan yang diterapkan agar kondisi masyarakat tidak semakin memburuk. Hal ini, sangat penting untuk dipertimbangkan agar pemerintah tidak perlu memberlakukan kebijakan PPKM jika diramalkan tidak akan mengalami peningkatan jumlah kasus dalam kurun beberapa waktu kedepan.

Secara umum tulisan ini merujuk pada penelitian sebelumnya yang ditulis oleh (Siregar, Pangruruk, & Widjaja, 2022) yang berjudul “Perbandingan Berbagai Model Peramalan Indeks Harga Saham Gabungan (IHSG) Di Masa Pandemi Covid-19”, memperlihatkan bahwa model PROPHET memberikan hasil peramalan terbaik. Dengan demikian, hal serupa diimplementasikan dalam penelitian ini untuk membuktikan apakah model PROPHET juga dapat digunakan untuk meramalkan jumlah kasus COVID-19 di Indonesia.

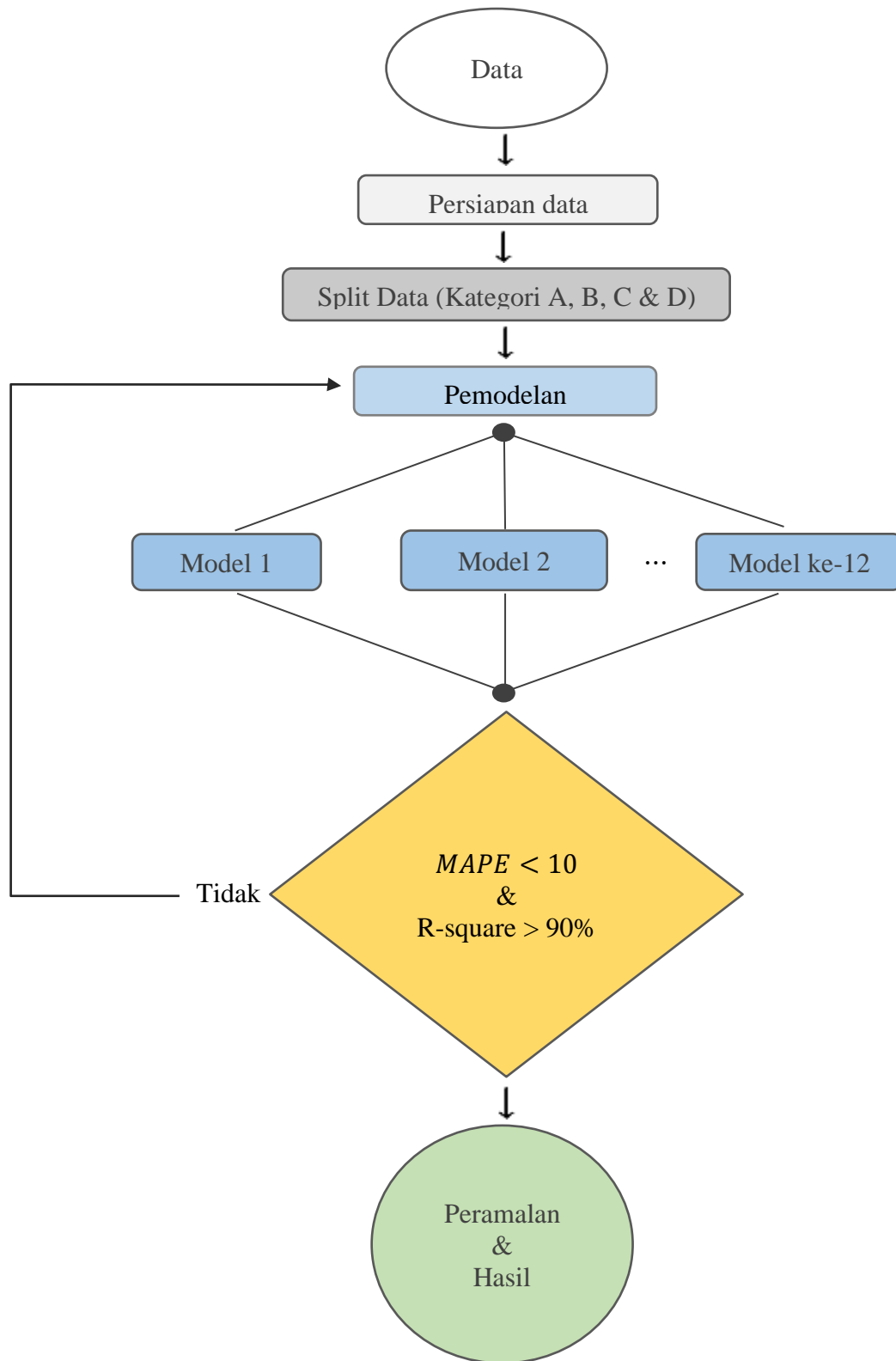
Penelitian ini menggunakan nilai MAPE sebagai tolak ukur ketepatan nilai dugaan model yang dinyatakan dalam bentuk rata-rata persentase absolut kesalahan dan lebih banyak digunakan untuk perbandingan pada data-data yang mempunyai skala interval waktu berbeda. Menurut (Putri, Zukhronah, & Pratiwi, 2021), Salah satu determinasi statistik yang digunakan untuk evaluasi setiap model runtun waktu yang dibandingkan adalah nilai Mean Absolute Percentage Error (MAPE). Dengan kata lain, hasil peramalan dikatakan semakin akurat jika nilai MAPE semakin kecil (Hudiyanti, Bachtiar, & Setiawan, 2019). Secara matematis dapat ditulis dengan $MAPE = \left(\frac{1}{n} \sum_{t=1}^n \left| \frac{y(t) - y'(t)}{y(t)} \right| \right) \times 100$, dimana $y(t)$ adalah nilai observasi pada waktu ke- t , $y'(t)$ adalah nilai peramalan pada waktu ke- t dan n adalah banyaknya observasi. Menurut (Purnama & Hendarsin, 2020) model yang baik memiliki nilai MAPE sesuai kriteria pada Tabel 1. Selain itu, R-square digunakan untuk memastikan seberapa cocok model regresi sesuai dengan data yang diamati. Jika hasil dari koefisien determinasi R-square di peroleh nilai yang sangat tinggi yaitu mendekati 1, dapat dikatakan model tersebut sudah sesuai dengan data yang diamati. Sebaliknya, semakin kecil nilai koefisien determinasi R-square, maka ini artinya pengaruh variabel bebas terhadap variabel terikat semakin lemah (Aryani, 2020) Alasan yang meperkuat penggunaan kedua ukuran statistik MAPE dan R ini merujuk pada penelitian yang dilakukan oleh (Chicco, Warrens, & Jurman, 2021), menunjukkan bahwa koefisien determinasi (R-kuadrat) lebih informatif dan benar daripada SMAPE, dan tidak memiliki keterbatasan interpretasi MSE, RMSE, MAE, dan MAPE. Sehingga, peneliti tak ragu untuk menggunakan R-square ini sebagai metrik standar untuk mengevaluasi model yang digunakan dalam penelitian ini.

Tabel 1
Kriteria Eवलuasi Model berdasarkan MAPE

Nilai MAPE (%)	Kriteria
< 10	Kemampuan peramalan sangat baik
10 – 20	Kemampuan peramalan baik
20 – 50	Kemampuan peramalan cukup
> 50	Kemampuan peramalan buruk

Metode Penelitian

Penelitian ini adalah metode penelitian kuantitatif dengan pendekatan analisis runtun waktu. Peneliti melakukan pengumpulan data historis jumlah kasus terkonfirmasi positif COVID-19 dari awal munculnya di Indonesia, Maret 2020 s/d Juli 2022 yang diperoleh dari salah satu user GitHub bernama Zakeigo. Kemudian, data historis tersebut disimulasikan menjadi empat kategori pembagian data latih dan data uji seperti diperlihatkan pada tabel 3. Selanjutnya, dibangun model dua belas model time series untuk meramalkan jumlah kasus COVID-19 dikemudian hari (satu bulan). Pada tahapan akhir dilakukan proses seleksi model peramalan terbaik dengan merujuk kepada masing-masing nilai MAPE dan R-square keduabelas model tersebut. Seluruh tahapan dalam penelitian ini dilakukan dengan bantuan software R, seperti diperlihatkan pada diagram atau alur penelitian gambar 1.

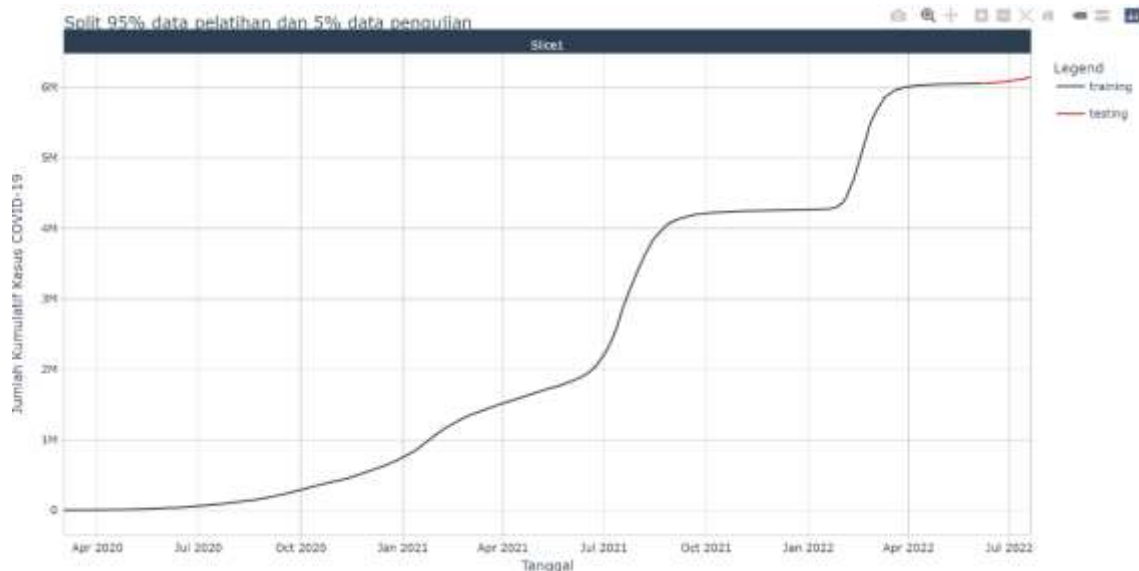


Gambar 1
Alur Metode Penelitian

Hasil dan Pembahasan

Berikut ini diperlihatkan gambaran umum jumlah kasus terkonfirmasi positif COVID-19 di Indonesia secara kumulatif pada gambar 2. Terlihat bahwa jumlah kumulatif kasus COVID-19 di Indonesia sudah tidak ada penambahan kasus yang signifikan sejak April 2022. Hal tersebut ter-indikasi vaksinasi sudah mencapai 97% untuk vaksinasi dosis 1, 81% dosis 2, dan juga vaksinasi dosis 3 sudah mencapai 24% (sumber; [Vaksin Dashboard \(kemkes.go.id\)](https://kemkes.go.id)). Tetapi, awal Juli 2022 terjadi peningkatan jumlah kasus terkonfirmasi positif dan ini menjadi perhatian penting bagi pemerintah. Kondisi ini juga menjadi landasan mendasar bagi peneliti merasa penting untuk melanjutkan penelitian ini. Dengan adanya model yang mampu memprediksi jumlah kumulatif kasus ini kedepannya diharapkan dapat digunakan menjadi rujukan bagi pemerintah dalam mengambil kebijakan dalam penanganan COVID-19 di Indonesia.

Jurnal ini hanya memperlihatkan simulasi proses penelitian menggunakan pembagian data kategori 4 (A, B, C, & D) seperti pada gambar 2, yaitu 95% data pelatihan yang digunakan untuk membangun berbagai model prediksi dan 5% sisanya.



Gambar 2
Total Kumulatif Jumlah Kasus Terkonfirmasi Positif COVID-19 Perhari

Data uji digunakan pada tahap evaluasi. Penelitian ini membandingkan beberapa model runtun waktu untuk memprediksi jumlah kumulatif COVID-19 seperti halnya yang dilakukan oleh (Duarte dkk, 2021), yang berjudul “Perbandingan prediksi deret waktu untuk indikator departemen darurat kesehatan dan dampak COVID-19”. Sehingga pada gambar 3 diperlihatkan hasil prediksi duabelas model yang telah dilakukan kalibrasi sesuai dengan masing-masing karakteristiknya, diperoleh bahwa NNET, KNN, RANGER (Random Forest), KERNLAB model no 9 (support vector machines) dan NAÏVE tidak melakukan prediksi dengan baik karena hasil prediksinya flat dan jauh dari tren data yang sebenarnya. Untuk memperkuat hasil pengamatan tersebut maka digunakan ukuran statistik yang dijadikan acuan evaluasi model yang diperlihatkan pada tabel 2. Jika merujuk pada nilai MAPE < 10 saja maka hanya ada dua model yang tidak dapat melakukan prediksi dengan baik yaitu; KNN dan Syntax Idea, Vol. 4, No. 8, Agustus 2022



Gambar 3
12 Model Runtun Waktu yang Digunakan Untuk Prediksi COVID-19

KERNLAB model no 9 (support vector machines). Jika ditinjau dari nilai R-square > 90%, diperoleh bahwa model NNET, KKNN, RANGER (Random Forest) dan NAÏVE adalah model yang tidak dapat merepresentasikan proporsi varians pada variabel dependen yang dapat dijelaskan oleh variabel independen variabel. Sehingga, dalam hal

Tabel 2
Evaluasi Model dengan MAE, MAPE, dll

No	Model	MAE	MAPE	MASE	SMAPE	RMSE	RSQ
1	ARIMA(0,2,2)	25126,51	0,41	11,82	0,41	33438,46	0,96
2	ETS(M,A,M)	5688,96	0,09	2,68	0,09	6469,18	0,97
3	LM	83050,27	1,37	39,06	1,38	103383,83	0,96
4	GLMNET	153141,57	2,52	72,03	2,57	192734,92	0,93
5	NNET	3785043,22	62,15	1780,24	90,17	3785135,41	0,00
6	KKNN	33280,96	0,54	15,65	0,55	42490,89	0,01
7	RANGER	33617,34	0,55	15,81	0,55	42754,87	0,00
8	KERNLAB Linear	85325,16	1,4	40,13	1,39	100420,04	0,96
9	SVM Nonlinear	941310,89	15,42	442,73	17,07	1047519,43	0,98
10	SNAIVE(7)	3410,61	0,56	16,08	0,56	43194,28	0,00
11	NNAR(1,1,10)[7]	92475,48	1,52	43,49	1,53	102458,13	0,58
12	PROPHET	84511,55	1,38	39,75	1,37	111000,24	0,95

Dengan merujuk pada nilai MAPE dan R-square yang dejalaskan sebelumnya maka dapat disimpulkan bahwa model terbaik yang dapat digunakan untuk memprediksi jumlah kumulatif terkonfirmasi positif COVID-19 di Indonesia terdapat tujuh model yaitu; ARIMA, ETS, LM, GLMNET, SVM Linear, NNAR, dan PROPHET yang diperlihatkan pada gambar 4. Jika harus memilih model terbaik dari terbaik maka adalah

ETS karena memiliki nilai determinasi statistik yang paling kecil pada MAE, MAPE, MASE, SMAPE, RMSE, dan memiliki nilai R-square tertinggi.

Secara keseluruhan percobaan yang dilakukan dalam penelitian ini dilampirkan pada tabel 3. Adapun model yang memiliki nilai R-square dan MAPE nya secara konsisten menjadi yang terbaik adalah PROPHET.

Tabel 3
Kategori Partisi Data Pelatihan dan Data Uji

No	Model	Persentasi Data Pelatihan Kategori							
		A	B	C	D	A	B	C	D
		Nilai R-Square				Nilai MAPE			
1	ARIMA(0,2,2)	0.54	0.83	0.80	0.95	08.44	12.83	00.22	00.43
2	ETS(M,A,M)	0.53	0.81	0.68	0.94	21.00	04.11	01.59	00.29
3	LM	0.54	0.83	0.80	0.95	13.22	11.29	04.93	01.20
4	GLMNET	0.80	0.61	0.87	0.91	13.57	09.75	05.75	02.10
5	NNET	0.00	0.00	0.00	0.00	71.01	68.96	65.27	61.97
6	KKNN	0.00	0.01	0.06	0.01	25.27	02.80	00.49	00.63
7	RANGER	0.00	0.00	0.00	0.00	25.36	03.11	00.49	00.64
8	KERNLAB Linear	0.54	0.83	0.80	0.95	13.55	10.61	04.56	01.49
9	SVM Nonlinear	0.82	0.70	0.70	0.97	57.50	35.77	35.06	19.89
10	SNAIVE(7)	0.00	0.00	0.00	0.00	25.60	03.51	00.51	00.65
11	NNAR(1,1,10)[7]	0.88	0.59	0.10	0.60	26.46	70.77	03.08	01.63
12	PROPHET	0.07	0.84	0.99	0.98	25.83	33.20	02.32	01.77

Kesimpulan

Sehubungan dengan tujuan dari penelitian ini adalah untuk menemukan model peramalan jumlah kasus COVID-19 di Indonesia dengan membandingkan duabelas model runtun waktu, diperoleh bahwa model terbaik dalam melakukan peramalan jumlah kasus COVID-19 tiga puluh hari mendatang (satu bulan) dapat digunakan model PROPHET. Selain itu, juga dibuktikan bahwa dalam memilih model terbaik tidak cukup menggunakan satu determinasi statistik saja, sangat dianjurkan untuk minimal menggunakan dua metrik yang ada (dalam hal ini R-square dan MAPE). Hal menarik lain yang ditemukan adalah penggunaan proporsi data pelatihan dan data uji juga sangat mempengaruhi akurasi prediksi model. Sehingga dapat disimpulkan bahwa ada beberapa model yang lebih unggul digunakan untuk prediksi jangka panjang tetapi tidak baik untuk prediksi jangka pendek. Adapun pengembangan model di masa mendatang, sangat disarankan untuk mencoba metodologi baru yang dapat digunakan untuk meramalkan kasus terkonfirmasi positif COVID-19 dengan mempertimbangkan variabel pengaruh lainnya.

BIBLIOGRAFI

- Aryani, Y. (2020). Sistem Informasi Penjualan Barang Dengan Metode Regresi Linear Berganda Dalam Prediksi Pendapatan Perusahaan. *Jurnal Riset Sistem Informasi Dan Teknologi Informasi (JURSISTEKNI)*, 39-51. [Google Scholar](#)
- Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*. [Google Scholar](#)
- Darmawansyah, D., Indar, I., Arifin, M. A., & Balqis, B. (2021). Evaluasi Program Penanggulangan Covid19 di Kota Palopo. *Jurnal Manajemen Kesehatan Yayasan RS. Dr. Soetomo*, 296-308. [Google Scholar](#)
- Diksa, I. G. B. N. (2021). Peramalan Gelombang Covid 19 Menggunakan Hybrid Nonlinear Regression Logistic–Double Exponential Smoothing di Indonesia dan Prancis. *Jambura Journal of Mathematics*, 37-51. [Google Scholar](#)
- Harini, S. (2020). Identification COVID-19 Cases in Indonesia with The Double Exponential Smoothing Method. *Jurnal Matematika MANTIK*, 66-75. [Google Scholar](#)
- Hudiyanti, C. V., Bachtiar, F. A., & Setiawan, B. D. (2019). Perbandingan Double Moving Average dan Double Exponential Smoothing untuk Peramalan Jumlah Kedatangan Wisatawan Mancanegara di Bandara Ngurah Rai. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2548. [Google Scholar](#)
- Kirbas, I. (2018). NAR based forecasting interface for time series analysis: T-seer. *International Conference on Engineering and Natural Science (ICENS)*, (pp. 44-149). [Google Scholar](#)
- Mawar, M., Andriyani, L., Gultom, A., & Ketiara, K. (2021). Dampak Sosial Ekonomi Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) di Indonesia. *Prosiding Seminar Nasional Penelitian LPPM UMJ* (p. 10623). Ciputat Jakarta Selatan: SEMNASLIT. [Google Scholar](#)
- Pakan, P. D. (2020). Peramalan Kasus Positif Covid 19 Di Indonesia Menggunakan LSTM. *Jurnal Ilmiah Flash*, 12-15. [Google Scholar](#)
- Purnama, D. I., & Hendarsin, O. P. (2020). Peramalan Jumlah Penumpang Berangkat Melalui Transportasi Udara di Sulawesi Tengah Menggunakan Support Vector Regression (SVR). *Jambura Journal of Mathematics*, 49-59. [Google Scholar](#)
- Putri, F. T., Zukhronah, E., & Pratiwi, H. (2021). Model ARIMA-GARCH Pada Peramalan Harga Saham PT. Jasa Marga (Persero). *Business Innovation and Entrepreneurship Journal*, 164-170. [Google Scholar](#)

Satrio, C. B., Darmawan, W., Nadia, B. U., & Hanafiah, N. (2021). Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET. *Procedia Computer Science*, 524-532. [Google Scholar](#)

Sayfeddine, D. (2014). Nonlinear autoregressive neural network with exogenous inputs based solution for local minimum problem of agent tracking using quadrotor. *Electronics Scientific Journal*, 2355. [Google Scholar](#)

Siregar, B., Pangruruk, F. A., & Widjaja, P. A. (2022). Perbandingan Berbagai Model Peramalan Indeks Harga Saham Gabungan (IHSG) di Masa Pandemi Covid-19. *Jurnal Multidisiplin Madani*, 1035–1046. [Google Scholar](#)

Wang, P., Zheng, X., Li, J., & Zhu, B. (2020). Prediction of Epidemic Trends in COVID-19 with Logistic Model and Machine Learning Technics. *Chaos, Solitons & Fractals*, 110058. [Google Scholar](#)

Copyright holder:

Bakti Siregar, F. Anthon Pangruruk, dan Simon Prananta Barus (2022)

First publication right:

[Syntax Idea](#)

This article is licensed under:

