

Model *Time Series* untuk Prediksi Jumlah Kasus Infeksi Coronavirus (Covid-19) di Sulawesi Selatan

Asrirawan

Program Studi Matematika, FMIPA, Universitas Sulawesi Barat, asrirawan@unsulbar.ac.id

Andi Seppewali

Program Studi Matematika, FMIPA, Universitas Sulawesi Barat, andi.seppewali@unsulbar.ac.id

Nurul Fitriyani

Jurusan Matematika, FMIPA, Universitas Mataram, nurul.fitriyani@unram.ac.id

ABSTRAK, Since it was declared a pandemic outbreak, the COVID 19 virus has become one of the main focuses of countries in the world in efforts to prevent the spread of the virus, including Indonesia. The areas of greatest severity in Indonesia include Jakarta, East Java, West Java and South Sulawesi. South Sulawesi Province is recorded as the largest area exposed to the COVID 19 pandemic outside Java Island. Predicting the number of COVID 19 cases is an alternative in preventing the spread through making government policies based on predictive data. This article presents a predictive model for the number of COVID 19 cases based on the ARIMA, Holt Winters and Nonlinear Autoregressive Neural Network (NAR-NN) Model. The results of the analysis show that the ARIMA Model (1,1,1) has a better level of prediction accuracy than the HW and NAR-NN models based on the MAPE criteria. Meanwhile, for the RMSE, MAE and MPE criteria, the NAR-NN model is better than others.

Kata Kunci: ARIMA, COVID 19, Holt Winters, NAR-NN

1. PENDAHULUAN

Coronaviruses Disease 19 (COVID 19) merupakan penyakit ancaman bagi seluruh negara di dunia. Penyakit yang disebabkan oleh virus *novel severe acute respiratory syndrome coronavirus 2* (SARS-CoV-2) telah ditetapkan sebagai wabah pandemik sejak tanggal 4 Mei 2020 [1]. Sebagian besar negara telah melakukan berbagai macam upaya pencegahan penyebaran COVID 19 melalui isolasi diri (*lockdown*), *physical distancing*, *Social distancing* atau dengan cara lain termasuk Indonesia. Indonesia telah terkonfirmasi jumlah kasus positif COVID 19 sebanyak 70.736 kasus per tanggal 09 Juli 2020 dengan penambahan kasus sebanyak 2.657 kasus. Beberapa wilayah Indonesia yang terdampak cukup parah adalah Jakarta, Jawa Timur, Jawa Barat dan Sulawesi Selatan.

Prediksi COVID 19 dengan menggunakan metode-metode peramalan statistika seperti metode peramalan *time series* diharapkan mampu membantu pemerintah khususnya di Provinsi Sulawesi Selatan dalam pencegahan melalui penyediaan data prediksi tersebut. Data tersebut menjadi salah satu indikator dalam penentuan daerah-daerah yang menjadi prioritas dalam menentukan kebijakan seperti *social distancing*, *physical distancing*, Pembatasan sosial Berskala Besar (PSBB), *lockdown*. Artikel ini akan membandingkan prediksi data COVID 19 di wilayah Sulawesi Selatan menggunakan model *time series* yakni, model Autoregressive Integrated Moving Average (ARIMA), *Nonlinear Autoregressive* (NAR) *Neural Networks*, dan *Holt Winters Exponential Smoothing*. Pemodelan COVID 19 menggunakan ARIMA dan NAR telah dilakukan sebelumnya oleh Khan dan Gupta [2] dengan menggunakan data infeksi COVID 19 di India dan diperoleh hasil prediksi jumlah kasus sebanyak 1500 kasus per hari untuk 50 hari ke depannya. Perbandingan ARIMA, Holt Winters Exponential Smoothing dan GARCH untuk prediksi COVID 19 di Nigeria juga pernah dilakukan oleh Abdulmajeed, Adeleke dan Pepoola [3]. pada artikel tersebut menyebutkan bahwa sangat susah memprediksi COVID 19 pada skenario kehidupan yang nyata dikarenakan adanya variabel-variabel yang cukup signifikan berpengaruh variabel jumlah tes, *randomness*, intervensi, pematuhan tinggal di rumah, faktor epidemik, dan variabel-variabel lainnya.

2. METODE PENELITIAN

Model Time Series

Metode peramalan data time series merupakan metode yang digunakan untuk memprediksi data selanjutnya berdasarkan data-data sebelumnya. Pada umumnya, pengukuran diamati pada interval waktu, sehingga menghasilkan discrete-time. Terdapat beberapa alasan analisis deret waktu dilakukan, misalnya untuk mempelajari dan menggambarkan fluktuasi pergerakan dari waktu ke waktu, untuk meramalkan nilai pada beberapa waktu yang akan datang, ataupun untuk mengontrol suatu sistem [4]. Artikel ini membandingkan model time series yakni model ARIMA, Holt Winters Exponential Smoothing dan Nonlinear Autoregressive (NAR) Neural Networks.

Model ARIMA

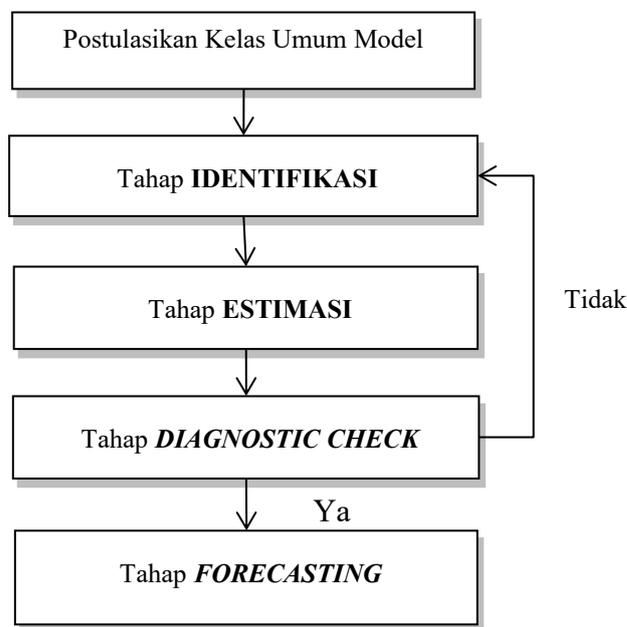
Beberapa model linier khusus untuk proses stasioner antara lain model autoregressive (AR), model moving average (MA), dan gabungan keduanya yaitu autoregressive moving average (ARMA). Model yang melibatkan ketidakstasioneran data Penjelasan mengenai model AR dan MA dapat dilihat pada Cryer [5] dan Wei [4]. Model ARIMA yang melibatkan komponen musiman biasa disebut dengan model SARIMA dimana pada waktu-waktu tertentu memiliki pola tersendiri dan pada umumnya terjadi pada lag-lag 12, 24, 36 dan seterusnya. Pemodelan ARIMA atau ARIMA Box-Jenkins sangat akurat pada peramalan data dalam jangka pendek sedangkan peramalan data jangka panjang kurang baik karena akan cenderung flat atau konstan. Adapun prosedur Box-Jenkins digunakan untuk memilih model ARIMA yang sesuai pada data deret waktu. Prosedur ini meliputi empat tahapan yang ditunjukkan oleh Gambar 1 Asrirawan [6], Haslina dkk [7].

Diberikan y_t adalah barisan deret waktu $\{y_1, y_2, \dots, y_t\}$, model ARIMA (p,d,q) dapat dilihat pada persamaan 1 [8].

$$\phi_p(B)(1-B)^d y_t = \theta_q(B)x_t$$

dengan $x_t \sim WN(0, \sigma^2)$, p orde AR, q orde MA, d orde pembeda, $\phi_p(B)$ dapat

dijabarkan dengan barisan suku $1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$, sedangkan $\theta_q(B)$ dapat ditulis $1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$. Penentuan orde p dan orde q dilihat dari cut off lag dari plot ACF dan PACF.



Gambar 1. Tahapan Prediksi Model ARIMA Box-Jenkins

Model Aditif Holt Winters: Exponential Smoothing dengan Trend dan Musiman

Model aditif Holt-Winters adalah pengembangan dari metode Holt Winters Exponential Smoothing dalam suatu rangkaian waktu metode peramalan untuk data univariat. penambahan faktor musiman ke ramalan trend, menjadi model data dengan trend yang sistematis atau komponen musiman. Metode ini masih sederhana dan merupakan metode peramalan yang kuat dan banyak digunakan karena dapat mengatasi variasi trend dan musiman dan dapat digunakan sebagai alternatif dari metode Box-Jenkins ARIMA yang populer. Namun, studi empiris cenderung menunjukkan bahwa metode ini tidak seakurat dengan model Box – Jenkins [1]. Model Holt Winters dapat dilihat pada persamaan 2 [8].

$$y_t = L_{t-1} + T_{t-1}$$

$$L_t = \alpha y_t + (1 - \alpha)(L_{t-1} + T_{t-1})$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

dengan α dan β adalah parameter penghalus.

Model Nonlinear Autoregressive (NAR) Neural Network

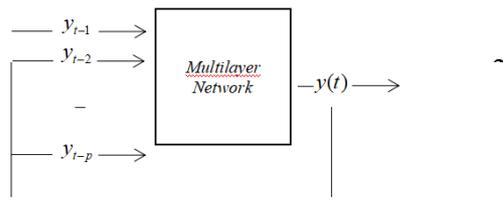
Model NAR-NN dapat digunakan untuk memprediksi data deret waktu pada waktu-waktu sebelumnya yang disebut dengan *feedback delays* $(y_{t-1}, y_{t-2}, \dots, y_{t-d})$ dengan d adalah parameter waktu tunda. Jaringan dibuat pada sebuah *open loop* yang menggunakan data real target sebagai variabel respon dan memastikan bahwa nilai tersebut mendekati nilai-nilai pada *training*. Setelah itu, langkah selanjutnya adalah jaringan di konversi ke *closed loop* dan nilai-nilai yang telah diprediksi digunakan untuk membuat variabel *input* respon yang baru di dalam jaringan. Model NAR-NN dapat ditulis pada persamaan 3.

$$y_t = h(y_{t-1}, y_{t-2}, \dots, y_{t-d}) + \varepsilon_t$$

dengan fungsi $h(\cdot)$ adalah fungsi yang tidak diketahui dan training pada jaringan digunakan untuk mendekati fungsi dari rerata optimasi dari bobot jaringan dan bias saraf. Model tersebut dapat diturunkan seperti pada persamaan 4.

$$y_t = \alpha_0 + \sum_{j=1}^k \alpha_j \phi \left(\sum_{i=1}^a \beta_{ij} + \beta_{0j} \right) + \varepsilon_t$$

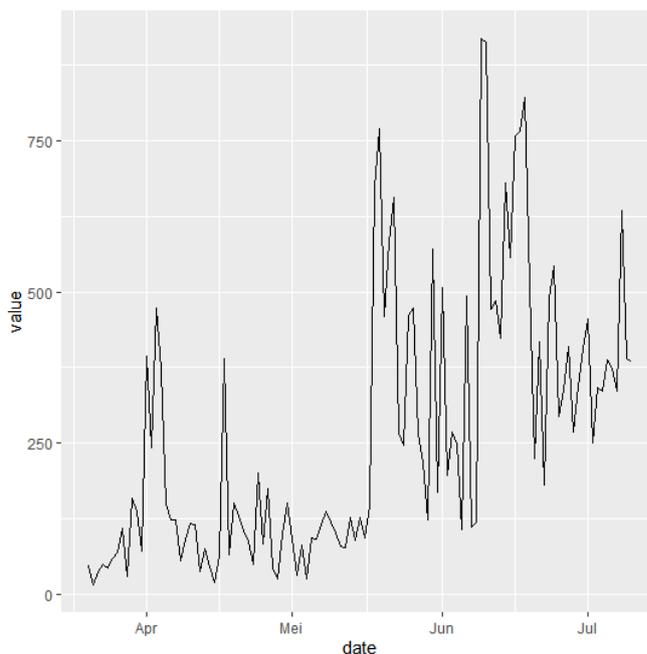
dimana a adalah jumlah entri, k adalah jumlah *hidden layers* dengan fungsi aktivasi ϕ , dan β_{ij} adalah parameter yang berkaitan dengan bobot antara koneksi input i dan *hidden layers* j sedangkan α_j adalah bobot antara koneksi *hidden layers* dan output. α_0 dan β_{0j} adalah konstanta *hidden layers* dan unit output [9]. Desain model NAR-NN dapat dilihat pada Gambar 2



Gambar 2 Desain Kerja NAR-NN

Data COVID 19 Sulawesi Selatan

Data harian infeksi COVID 19 Sulawesi Selatan diperoleh dari hasil publikasi Pemerintah Provinsi Sulawesi Selatan yang dapat di akses pada laman web <https://covid19.sulselprov.go.id/data>. Data tersebut tersedia mulai tanggal 20 Maret 2020 sampai dengan 10 Juli 2020 dengan tanggal penulsuran 11 Juli 2020. Jumlah data harian yang digunakan sebanyak 113 hari. Data kemudian dibagi kedalam data *training* sebanyak 89 hari dan data testing sebanyak 24 hari. Penyebaran kasus positif di Sulawesi Selatan yang terbagi atas 24 kabupaten kota dapat dilihat pada Gambar 3. Selain itu, data dianalisis menggunakan program R.



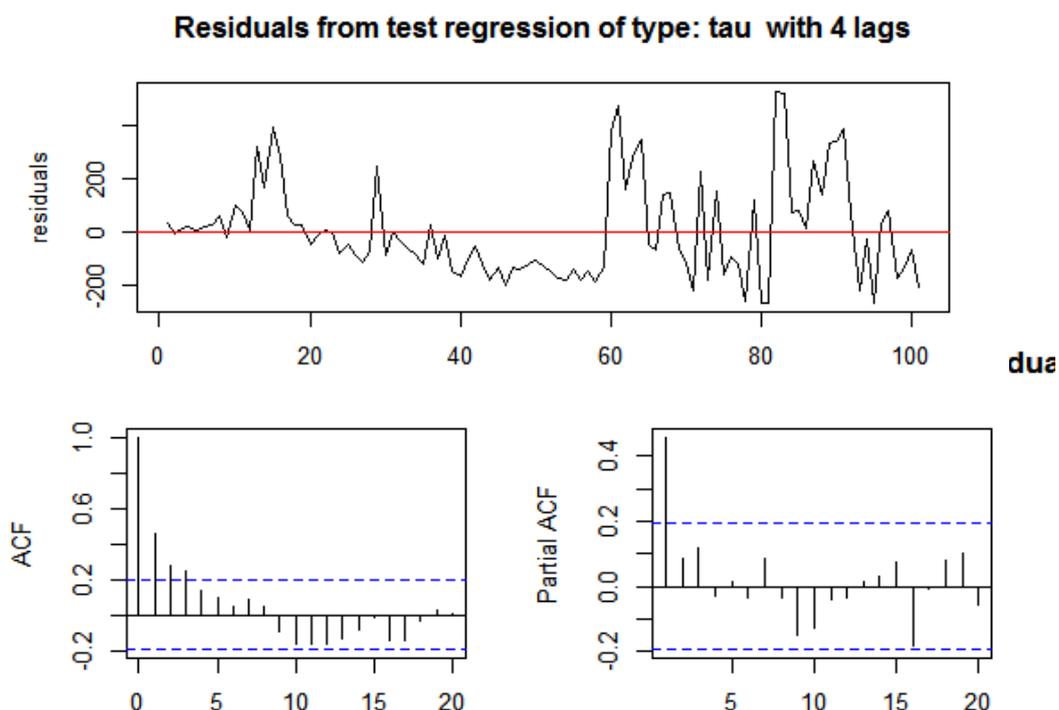
Gambar 3. Jumlah Kasus Positif COVID 19 Sulawesi Selatan (Maret-Juli)

3. PEMBAHASAN

Evaluasi Model ARIMA, Holt Winters NAR-NN

Data harian positif COVID 19 Sulawesi Selatan mengalami pergerakan cenderung naik atau trend naikan dan terlihat bahwa data tidak stasioner dalam rata-rata sehingga dilakukan differencing. Hasil differencing kedua dapat dilihat pada Gambar 4. Berdasarkan Gambar 4 bahwa data terlihat sudah stasioner dengan plot ACF dan PACF. Berdasarkan plot-plot tersebut diperoleh model ARIMA (1,1,1) yang selanjutnya dilakukan estimasi menggunakan MLE (forecast dan fUnitRoots package di R) diperoleh model ARIMA terbaik yakni ARIMA (1,1,1) dengan nilai AIC dan BIC masing-masing 1314,27 dan 1322,08.

tersebut, model ARIMA (1,1,1) telah memenuhi distribusi normal residual dengan nilai p-value > 0,05 dan nilai uji lag-lag 12, 24, 36, dan lag seterusnya telah memenuhi uji white noise (p-value > 0,05). Evaluasi kriteria kebaikan model dan prediksi model ARIMA disajikan pada tabel 1. Akurasi pemodelan additive Holt Winters yang melibatkan trend pada data COVID 19 dapat dilihat pada tabel 1 dengan parameter penghalus α sebesar 0,4023 dan β sebesar 0,0001. Nilai AIC dan BIC untuk model HW masing-masing sebesar 1512,62 dan 1525,69. Nilai tersebut terlihat lebih besar dibandingkan dengan model ARIMA (1,1,1) pada data *training*. Sedangkan untuk Model NAR-NN menggunakan neuron input, output dan layer tersembunyi masing-masing satu sedangkan neuron dalam layer yang tersembunyi sebanyak tiga.



Gambar 4. Time series plot, ACF dan PACF setelah differencing kedua

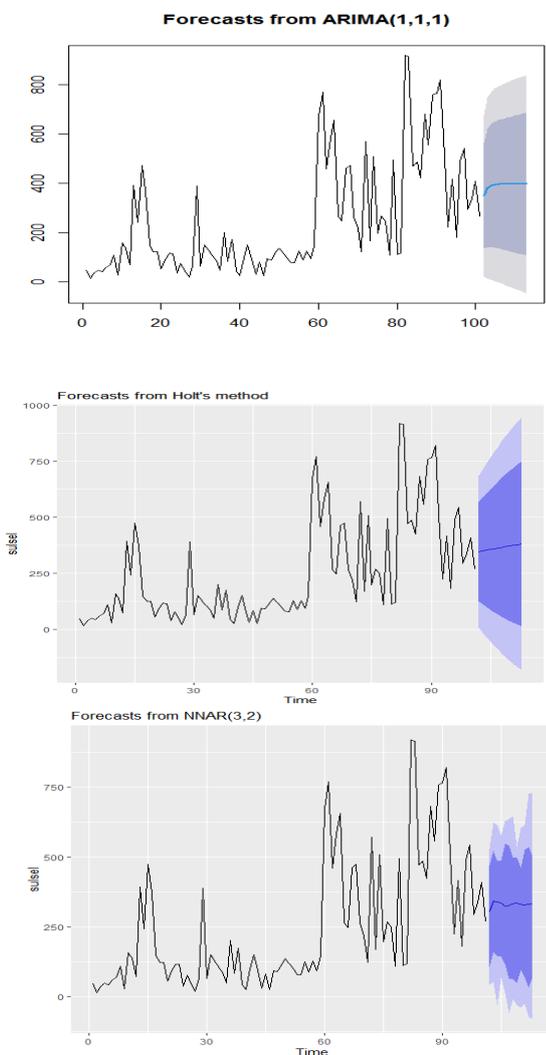
Nilai estimasi parameter AR dan MA dengan menggunakan ML pada model ARIMA (1,1,1) adalah 0,3647 dan -0,8592 (*lmtest package*). Uji diagnostik telah dilakukan dengan menguji kenormalan residual dan white noise menggunakan LjungBox pada program r. Dari uji

Perbandingan model terbaik untuk data training pada data COVID 19 Sulawesi Selatan menggunakan empat kriteria yakni, RMSE, MAE, MPE, dan MAPE. Nilai akurasi ketiga model telah disajikan pada tabel 1.

Tabel 1. Tingkat Akurasi Kebaikan *Model Time Series*

Model	RMSE	MAE	MPE	MAPE
ARIMA (1,1,1)	166,29	110,49	-32,30	63,61
Holt Winters	169,21	114,18	-37,79	65,14
NAR-NN	138,10	98,86	-47,79	67,78

Berdasarkan tabel 1, akurasi model time series untuk data COVID 19 Sulawesi Selatan menunjukkan bahwa model NAR-NN memiliki tingkat akurasi prediksi yang lebih dibandingkan dengan model lainnya untuk kriteria RMSE, MAE, dan MPE sedangkan untuk kriteia MAPE menunjukkan bahwa model ARIMA (1,1,1) lebih baik dibandingkan dengan model lainnya. Hasil plot peramalan untuk masing-masing model dapat dilihat pada Gambar 5.



Gambar 5. Prediksi Model 12 Hari kedepan

4. KESIMPULAN

Prediksi data harian COVID 19 untuk wilayah Sulawesi Selatan dengan menggunakan model ARIMA, Holt Winters dan NAR-NN memiliki tingkat akurasi yang cukup besar hal ini disebabkan oleh beberapa faktor seperti jumlah data deret waktu yang masih terbatas, epidemik, randomness dan lain-lain. Berdasarkan kriteria RMSE, MAE dan MPE, model NAR-NN memiliki tingkat akurasi prediksi yang lebih baik dibandingkan dengan model ARIMA dan HW. Sedangkan untuk kriteria MAPE, model ARIMA (1,1,1) lebih baik dibandingkan dengan model lainnya.

5. DAFTAR PUSTAKA

- [1] V. Papastefanopoulos, P. Linardatos, and S. Kotsiantis, 2020, "COVID-19: A Comparison of Time Series Methods to Forecast Percentage of Active Cases per Population", *Appl. Sci.* 2020, 10(11), 3880; <https://doi.org/10.3390/app10113880>, 2020.
- [2] M. F. Khan, and R. Gupta, "ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India", *Journ. Safety Science and Resilience*, 2020, 1(12), <https://doi.org/10.1016/j.jnlssr.2020.06.007>. 2020.
- [3] K. Abdulmajeed, M. Adelekeb, and L. Popoola, "Online Forecasting of COVID-19 Cases in Nigeria using Limited Data", *Data in Brief*, 30, <https://doi.org/10.1016/j.dib.2020.105683>. 2020.
- [4] W.W.S. Wei, "Time Series Analysis, Univariate and Multivariate Methods", 2nd ed. United States: Addison-Wesley Publishing Company. 2006.
- [5] J. D. Cryer, "Time Series Analysis", Boston: PWS-KENT Publishing Company. 1986.
- [6] Asrirawan, "Simulasi Perbandingan Metode Peramalan Model Generalized Seasonal Autoregressive Integrated Moving Average (GSARIMA) Dengan

- Seasonal Autoregressive Integrated Moving Average (SARIMA)”, *Jurn. Dinamika*, 1(6), Universitas Cokroaminoto Palopo. 2015.
- [7] Haslina, Hasmah, K. W. Fitriani, M. Asbar, dan Asrirawan, “Penerapan Metode ARIMA (Autoregressive Integrated Moving Average) Box Jenkins Untuk Memprediksi Pertambahan Jumlah Penduduk Tansmigran (Jawa Dan Bali) Di Kecamatan Sukamaju, Kabupaten Luwu Utara Propinsi Sulawesi Selatan”, *Jurn.Dinamika*, 1(9), Universitas Cokroaminoto Palopo. 2018.
- [8] P. J. Brockwell and R. A. Davis, “Introduction to Time Series and Forecasting, 3rd ed”. New York: Springer-Verlag New York. 2016
- [9] G. Benrhmach, Namir, A. Namir and J. Bouyaghroumni, “Nonlinear Autoregressive Neural Network and Extended Kalman Filters for Prediction of Financial Time Series”, *Journal of Applied Mathematics* Volume 2020 (6), <https://doi.org/10.1155/2020/5057801>. 2020