



Prediksi Penyebaran Demam Berdarah Dangué dengan Algoritma Hybrid Autoregressive Integrated Moving Average dan Artificial Neural Network: Studi Kasus di Kabupaten Bandung

**Ichwanul Muslim Karo*

Fakultas Informatika Universitas Telkom
Bandung, Indonesia

*Correspondence: E-mail: ichwanulkarokaro@telkomuniversity.ac.id

ABSTRACT

Demam Berdarah Dengue (DBD) merupakan penyakit menular yang ditularkan melalui gigitan nyamuk *Aedes Aegypti*. WHO (World Health Organization) telah mengupayakan langkah-langkah pencegahan terhadap wabah DBD dengan penerapan teknologi. Teknologi yang digunakan untuk mencegah penyebaran wabah DBD adalah penggunaan serangkaian proses komputasi untuk menghasilkan prediksi penyebaran DBD yang diharapkan dapat membantu langkah pencegahan. Dalam membantu pengembangan teknologi pencegahan DBD penulis mengembangkan model hybrid Autoregressive Integrated Moving Average (ARIMA) dan Artificial Neural Network (ANN) untuk membantu memprediksi incident rate DBD berdasarkan beberapa variabel terkait seperti cuaca dan incident rate yang diambil dari Januari 2009 – November 2016. Dari model hybrid ARIMA dan ANN dihasilkan nilai prediksi yang memiliki tingkat error yang rendah yang diindikasikan oleh nilai RMSE yang kecil. Model hybrid ARIMA-ANN yang optimal adalah hybrid ARIMA-ANN dengan orde (1,0,3) dengan nilai RMSE sebesar 0.0087.

ARTICLE INFO

Article History:

Submitted/Received 02 Maret 2021

First Revised 20 Apr 2021

Accepted 01 May 2021

First Available online 20 Mai 2021

Publication Date 01 Juni 2021

Keyword:

ARIMA,

ANN,

DBD,

Incident rate.

1. PENDAHULUAN

Demam Berdarah Dengue (DBD) adalah penyakit menular via gigitan nyamuk *Aedes Aegypti*. Gigitan nyamuk mengandung virus *Flaviviridae* yang dapat kematian bagi penderita. Penyebaran penyakit DBD di area perkotaan lebih pesat dibandingkan di area pedesaan karena faktor tingginya kepadatan penduduk (Cholil, *et al.*, 2020). Selain itu Penyebaran DBD dipengaruhi oleh factor cuaca dan iklim, akan meningkat jika terjadi musim hujan. Dinas Kesehatan Kabupaten Bandung mencatat bahwa selain covid-19, penyakit DBD masih tinggi dan perlu diwaspadai penyebarannya. Data Dinkes Kabupaten Bandung mencatat bahwa terjadi kenaikan kasus DBD di wilayah Kabupaten Bandung dalam rentang Agustus 2018 sampai Januari 2019, pada Agustus 2018 terjadi 28 kasus, September 135 kasus, Oktober 159 kasus, November 170 kasus, Desember 148 kasus dan sampai dengan 29 Januari terjadi 236 kasus (Fuadzy, *et al.*, 2021).

Salah satu upaya yang dilakukan untuk menekan penyebaran DBD adalah dengan melakukan prediksi terhadap penyebarannya, sehingga diperoleh insight bagi Pemerintah untuk menetapkan kebijakan (Ichwani, *et al.*, 2019). Penerapan prediksi penyebaran DBD telah banyak dilakukan untuk berbagai daerah dengan mempertimbangkan unsur iklim dan cuaca (Kristianto, *et al.*, 2017)

Salah satu metode yang dapat digunakan untuk melakukan proses prediksi data adalah Auto Regressive Integrated Moving Average (ARIMA) dan metode kecerdasan buatan seperti Artificial Neural Network (ANN). Hybrid Arima-Ann sudah pernah diimplementasikan untuk memprediksi DBD di Kabupaten Sragen (Setiawan, *et al.*, 2020) Metode ARIMA sering digunakan sebagai metode standar untuk melakukan prediksi data timeseries (Satrio, *et al.*, 2021). Faktanya bahwa setiap kejadian tidak mungkin berdiri sendiri, tetapi juga dipengaruhi oleh variabel lain sehingga metode ARIMA tidak baik diteapkan pada data dengan pola campuran ini (Büyüksahin, *et al.*, 2019). Kekurang pada Arima diatasi dengan metode Artificial Neural Network (ANN). Kelebihan jaringan ANN adalah pada control area, prediksi, dan pengenalan pola. Sehingga luaran yang dihasilkan mampu mendekati nilai sebenarnya (Setiawan, *et al.*, 2020).

Terdapat beberapa penelitian sebelumnya yang digunakan sebagai rujukan. Kristianto, dkk. memprediksi penderita DBD di Kabupaten Sragen menggunakan algoritma Triple Exponential Smoothing (TES) yakni variasi dari model Arima (Kristianto, *et al.*, 2017). Mereka menggunakan data penderita DBD dari Juni 2013 hingga 2016. Jenis Data yang digunakan adalah timeseries. Model evaluasi yang digunakan adalah MAPE dengan tingkat error 9 persen (Nabilah dan Ranggadara, 2020). Pada area studi kasus yang sama, Setiawan dkk mengkaji dan menganalisis penggunaan kombinasi metode Arima dan ANN dengan Estimasi Titik dan Interval untuk meningkatkan akurasi prediksi penderita demam berdarah dengue (DBD).

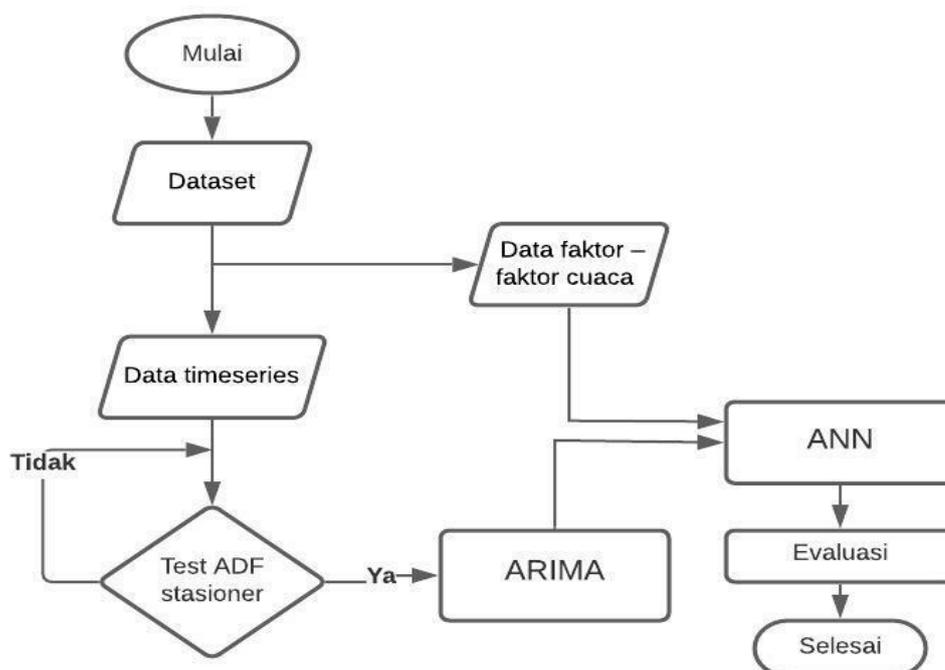
Dompas, dkk juga memprediksi penderita DBD di Kepulauan Sangihe menggunakan metode Arima (Dompas, et al., 2021). Data yang digunakan adalah jumlah penderita dalam kurun 2011 hingga 2019. Adapun dalam percobaannya, mereka memprediksi angka penderita DBD lima tahun mendatang dengan menggunakan berbagai scenario parameter.

Kombinasi Arima-Ann dinilai cukup efektif untuk memprediksi berbagai variabel eksternal yang mempengaruhi penyebaran DBD, seperti cuaca (Team, et al., 2009) , temperatur (Ammar, et al., 2019), dan curah hujan (Dwivedi, et al., 2019). Sehingga Kombinasi Arima-Ann dinilai sangat efektif untuk memprediksi penyebaran DBD di Kabupaten Bandung

Berdasarkan uraian di atas, penulis tertarik untuk mengkaji dan menganalisis ulang penggunaan kombinasi metode Auto Regressive Integrated Moving Average(ARIMA) dan Artificial Neural Network(ANN) dengan Data IR (Incident Rate) DBD dan data cuaca (penyinaran, kelembapan, angin, temperatur, uap, dan curah hujan) di Kabupaten Bandung.

2. METODE PENELITIAN

Penelitian ini menggunakan metode hyrid ARIMA – ANN. Penelitian ini memiliki dua tahap yakni proses peramalan menggunakan metode ARIMA dan dilanjutkan dengan prediksi menggunakan ANN (Gambar 1). Pada tahap pertama dilakukan identifikasi model ARIMA dan menentukan nilai parameter model ARIMA menggunakan ADF (Augmented Dickey Fuller) (Cheung and Lai, 2019). Setelah didapat nilai Stasioner dari ADF (Augmented Dickey Fuller) maka model ARIMA bisa dijalankan dan menghasilkan nilai prediksi. Proses terakhir yakni evaluasi hasil prediksi menggunakan RMSE.



Gambar 1 Kerangka Kerja Penelitian

2.1. Dataset

Dalam penelitian ini terdapat dua jenis dataset, yakni data Incident Rate (IR) dan Data faktor cuaca. Jenis data IR adalah data timeseries. Data IR diperoleh dari Dinas Kesehatan Kabupaten Bandung dari tahun 2009 hingga 2016. Incident Rate atau rata – rata kejadian adalah nilai yang didapat dengan menghitung jumlah kasus demam berdarah pada bulan tertentu di suatu daerah berpopulasi tertentu (Kristianto, *et al.*, 2017). Secara matematis dapat dihitung dengan menggunakan persamaan (Cholil, *et al.*, 2020).

Dimana case adalah jumlah kasus kejadian yang terjadi di suatu daera

$$R = \frac{\text{case} \times 1000}{\text{populasi}} \quad (\text{Cholil, et al., 2020})$$

2.2 ARIMA

ARIMA (Autoregressive Integrated Moving Average) adalah metode yang digunakan untuk memprediksi kejadian diwaktu mendatang dengan menggunakan data dari waktu sebelumnya (Rasmanto, *et al.*, 2017). Persamaan umum Arima dinyatakan dengan ARIMA (Bandung, *et al.*, 2019).

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^q \phi_{jqi} e_{t-j} + \varepsilon_t \quad (\text{Bandung, et al., 2019})$$

Pada penelitian ini,, model Arima yang digunakan ditambahkan variabel eksogenus, sehingga diperoleh sebagai

$$y_t = \beta x_t + \mu + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^q \phi_{jqi} e_{t-j} + \varepsilon_t \quad (\text{Ichwani, et al., 2019})$$

Dimana p adalah orde autoregressive, $yt-i$ adalah data pada saat waktu (t) tertentu, q adalah orde Moving Average, $\epsilon t-j$ adalah error pada saat t tertentu, d adalah differencing, θ adalah konstanta, β adalah parameter eksogenus dan μ adalah intercept.

Dalam metode ARIMA terdapat istilah ACF (Autocorrelation Function) yang digunakan untuk menentukan orde q (Ching and Phoon, 2019). Selain fungsi ACF, terdapat fungsi PACF (Partial Autocorrelation Function) yang berfungsi untuk menentukan orde dari q (Ching and Phoon, 2019). Fungsi ACF dan PACF ini dipanggil menggunakan library yang ada pada bahasa pemrograman Python.

Untuk menentukan parameter pada model ARIMA digunakan fungsi MLE (Maximum Likelihood) untuk memaksimalkan fungsi peluang pada ARIMA. Selain ACF untuk menentukan orde q dan PACF untuk menentukan orde p . Terdapat uji ADFuller (Augmented Dickey – Fuller) untuk menguji stasioneritas data. Dalam uji ADFuller diuji 2 hipotesis yaitu:

1. H_0 = Data time series tidak stasioner
2. H_1 = Data time series stasioner

Apabila data tidak stasioner maka perlu dilakukan proses differencing sampai data menjadi stasioner. mana persamaan (Ichwani, et al., 2019).

2.3 ANN

Metode ANN (Artificial Neural Network) terinspirasi dari jaringan syaraf manusia (Jani, et al., 2017).. Terdapat tiga bagian dalam ANN yaitu input, hidden layer, dan output. Pada penelitian ini digunakan ANN dengan fungsi aktivasi linear (persamaan 4)

$$f(x) = x \quad (\text{Kristianto, et al., 2017}).$$

Dengan ketengan x adalah jumlah *input*, karena nilai $f(x)$ adalah x itu sendiri maka tidak diperlukan normalisasi pada hasil datanya. Hal ini disebabkan *output* dari fungsi aktivasi tersebut sudah merupakan hasil prediksi.

2.4. EVALUASI MODEL

Untuk mengestimasi tingkat akurasi hasil prediksi digunakan fungsi RMSE (Root Mean Square Error) menggunakan persamaan (Rasmanto, et al., 2017).

$$RMSE = \frac{\sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2}}{n} \quad (\text{Rasmanto, et al., 2017})$$

Keterangan:

n = jumlah incident rate
 \hat{y}_i = prediksi incident rate saat i
 y_i = data incident rate saat i

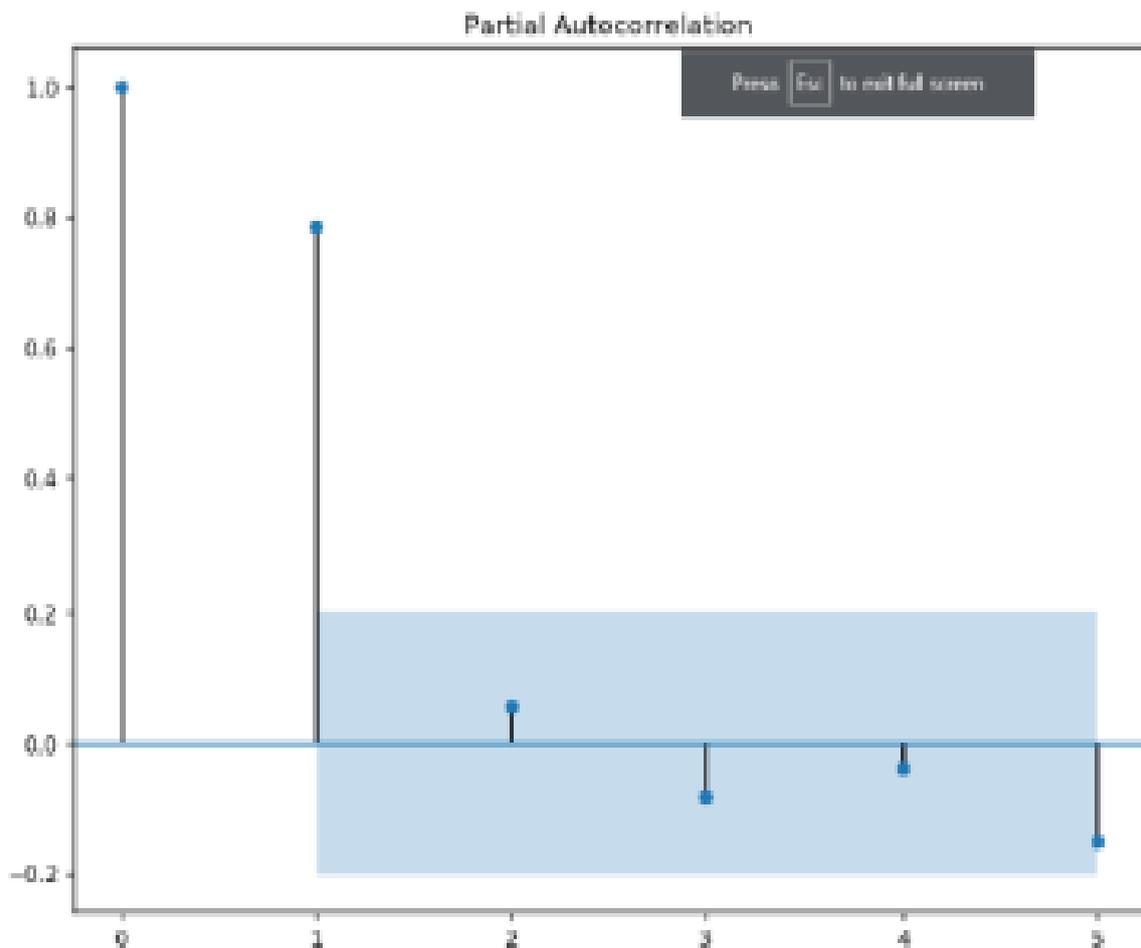
3. HASIL DAN PEMBAHASAN

Tahap pertama menentukan orde dari ARIMA yang akan digunakan dengan melakukan pengujian ADFuller dan pemeriksaan plot ACF dan PACF pada data incident rate. Uji ADFuller menguji 2 hipotesis yaitu:

1. H_0 berarti data time series bersifat tidak stasioner
2. berarti data time series bersifat stasioner

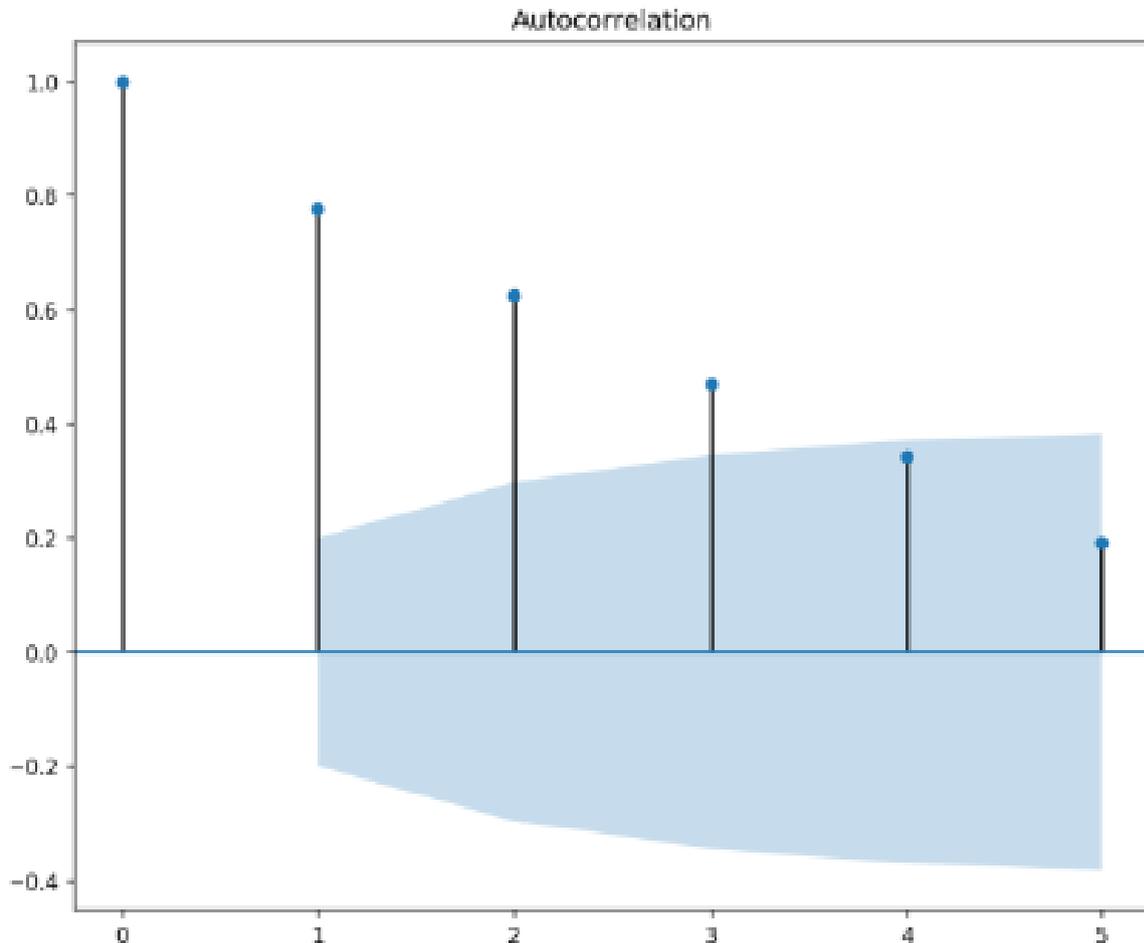
Pada uji ADFuller yang dilakukan, p-value dari data incident rate adalah 0,005928 dimana nilai ini lebih

kecil dari 0,05 sehingga data tidak perlu melalui proses differencing. Untuk menentukan orde ARIMA maka dilakukan pembacaan pada plot ACF dan PACF (Gambar 2 dan 3).



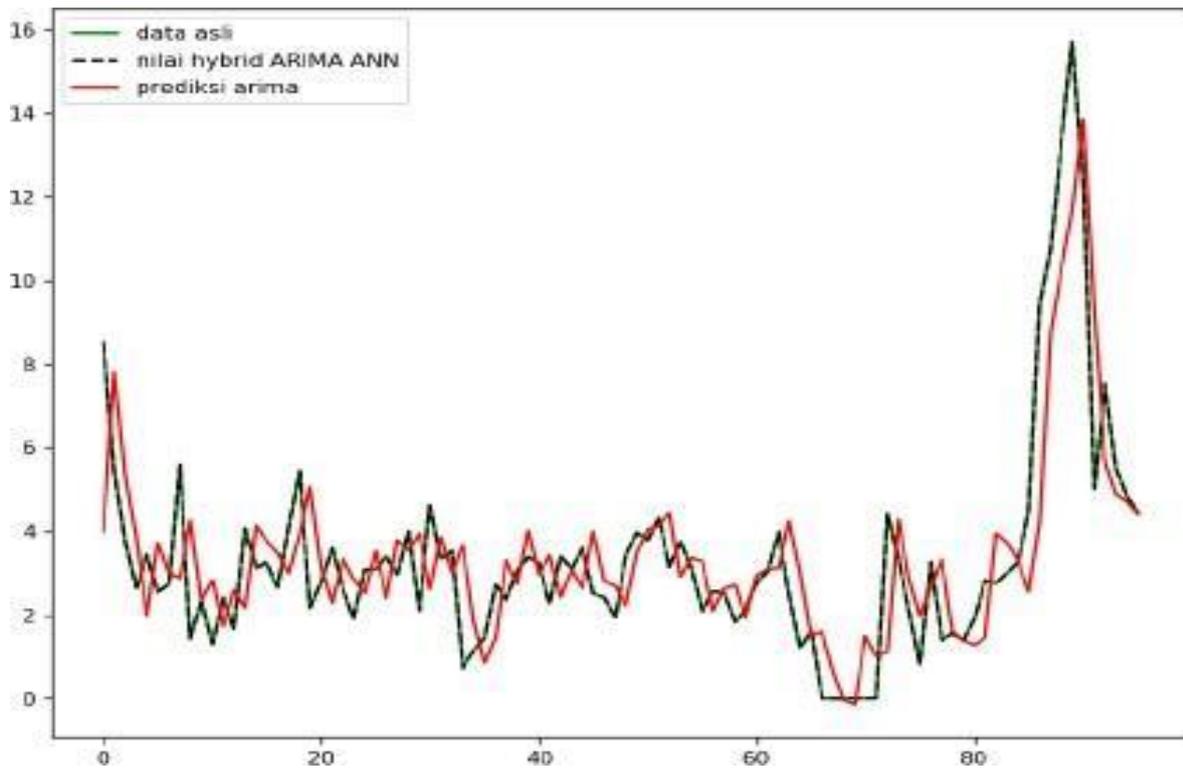
Gambar 2 Grafik Plot PACF

Gambar 2. Merupakan visualisasi plot PACF, daerah warna biru menggambarkan tingkat signifikansi. Gambar 3 menunjukkan orde p (Autoregressive) bernilai 1 karena lag 1 merupakan lag terakhir yang tidak berada dalam daerah tingkat signifikansi. Gambar 3 merupakan visualisasi plot ACF dengan daerah warna biru menggambarkan tingkat signifikansi. Gambar 2 menunjukkan orde q (Moving Average) bernilai 3 karena lag 3 merupakan lag terakhir yang tidak berada dalam daerah tingkat signifikansi.



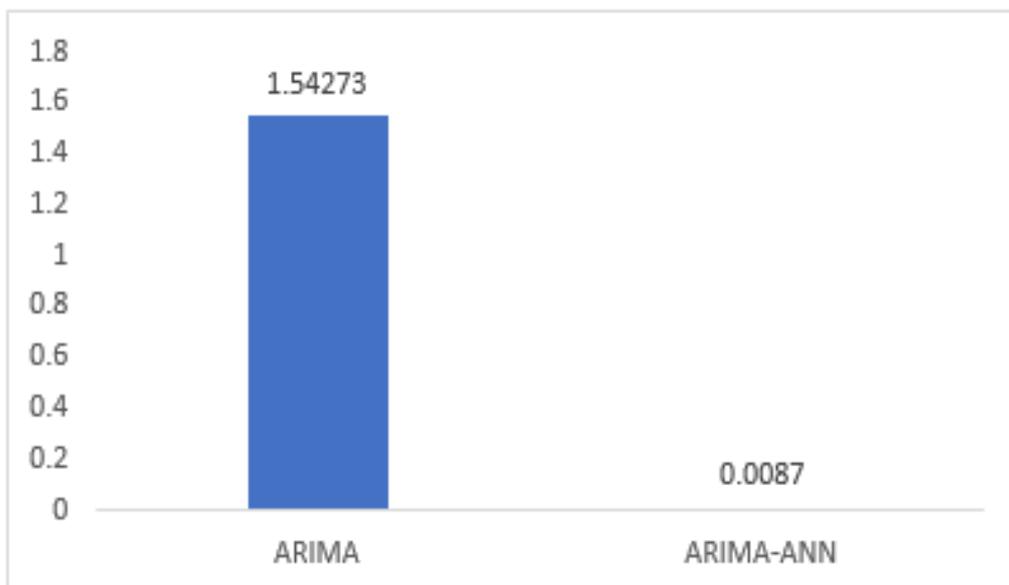
Gambar 3 Grafik Plot ACF

Gambar 4 adalah visualisasi dari nilai Incident Rate (IR) asli, IR hasil prediksi ARIMA, dan IR hasil prediksi ARIMA-ANN. Visualisasi diatas menunjukkan bahwa nilai IR hasil prediksi ARIMA-ANN memiliki grafik yang paling mirip dengan nilai IR asli dan dibuktikan dengan nilai RMSE ARIMA (1,0,1) – ANN sebesar 0.00846. Nilai ini lebih kecil dari RMSE ARIMA (1,0,1) yang bernilai 1.55475.



Gambar 4 Grafik Visualisasi ARIMA (1,0,1) dan ARIMA (1,0,1) – ANN

Berdasarkan hasil studi literasi yang dilakukan, ditemukan bahwa kejadian demam berdarah dipengaruhi oleh suhu udara (temperatur), kelembapan, curah hujan dan lamanya penyinaran matahari, sementara itu kecepatan angin tidak terlalu berpengaruh signifikan. Selanjutnya, berdasarkan hasil percobaan, nilai RMSE model ARIMA-ANN jauh lebih kecil dibandingkan dengan ARIMA (Gambar 5), dengan nilai MSE 0,00008 dan ADF Statistic = -3,591370. Artinya Metode ARIMA-ANN lebih baik dalam memprediksi penyebaran DBD di kabupaten Bandung dibandingkan model ARIMA.



Gambar 5 Perbandingan RMSE

4. CONCLUSION

Variabel cuaca berpengaruh terhadap incident rate demam berdarah di Kabupaten Bandung adalah suhu udara (temperatur), kelembapan, curah hujan dan lamanya penyinaran matahari, sementara itu kecepatan angin tidak terlalu berpengaruh signifikan. Model hybrid ARIMA - ANN memiliki hasil prediksi lebih baik dibandingkan dengan model ARIMA dengan nilai RMSE sebesar 0.00870.

5. REFERENCES

- Ammar, G., & Ammar, E. J. K. (2019). A Hybrid ARIMA and Artificial Neural Networks (ANN) Model for Daily Maximum and Minimum Temperature Forecast. *Tishreen University Journal-Engineering Sciences Series*, 41(6).
- Büyüksahin, Ü. Ç., dan Büyüksahin, E. Ş. (2019). Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition. *Neurocomputing*, 361, 151-163.
- Ching, J., & Phoon, K. K. (2019). Impact of autocorrelation function model on the probability of failure. *Journal of Engineering Mechanics*, 145(1), 144-160.
- Cheung, Y. W., & Lai, K. S. (1995). Lag order and critical values of the augmented Dickey–Fuller test. *Journal of Business & Economic Statistics*, 13(3), 277-280.
- Cholil, S. R. D. A. F., dan Cholil, A. T. (2020). Prediksi penyakit demam berdarah di Puskesmas Nemplak Simongan menggunakan algoritma C4.5. *SISTEMASI: Jurnal Sistem Informasi*, 9(3), 529-542.
- Dompas, J. V. K., dan Dompas, J. S. P. M. S. (2021). Prediksi jumlah penderita demam berdarah di Kabupaten Kepulauan Sangihe menggunakan model ARIMA musiman. *d'CARTESIAN*, 9(2), 126-132.
- Dwivedi, D. K., Dwivedi, J. H., dan Dwivedi, S. G. R. (2019). Forecasting monthly rainfall using autoregressive integrated moving average model (ARIMA) and artificial neural network (ANN) model: A case study of Junagadh, Gujarat, India. *Journal of Applied and Natural Science*, 11(1), 35-41.
- Fuadzy, H., Prasetyowati, H., Marliyanih, E. S., Hendra, A., & Dadang, A. M. (2021). Autokorelasi Spasial Demam Berdarah Dengue di Kota Tasikmalaya. *ASPIRATOR-Journal of Vector-borne Disease Studies*, 13(2), 113-126.
- Ichwani, A. S., dan Ichwani, W. H. A. (2019). Prediksi Angka Kejadian Demam Berdarah Dengue (DBD) Berdasarkan Faktor Cuaca Menggunakan Metode Extreme Learning Machine

(Studi Kasus Kecamatan Tembalang). *Jurnal Iptekmedia Komunikasi Teknologi*, 3(1), 31-24.

Jani, D. B., Jani, M. M., dan Jani, S. P. K. (2017). Application of artificial neural network for predicting performance of solid desiccant cooling systems—A review. *Renewable and Sustainable Energy Reviews*, 80, 352-366.

Kristianto, R. P. U. E., dan Kristianto, L. E. T. (2017). Penerapan Algoritma Forecasting untuk Prediksi Penderita Demam Berdarah Dengué di Kabupaten Sragen. *Semnasteknomedia Online*, 5(1).

Karyanti, M. R., dan Hadinegoro, S. R. (2016). Perubahan epidemiologi demam berdarah dengué di Indonesia. *Sari Pediatri*, 10(6), 424-432.

Kusuma, A. P., dan Sukendra, D. M. (2016). Analisis spasial kejadian demam berdarah dengué berdasarkan kepadatan penduduk. *Unnes Journal of Public Health*, 5(1), 48-56.

Nabillah, I., & Ranggadara, I. (2020). Mean Absolute Percentage Error untuk Evaluasi Hasil Prediksi Komoditas Laut. *Journal of Information System*, 5(2), 250-255.

Perwitasari, D., dan Perwitasari, A. Y. (2015). Model prediksi demam berdarah dengué dengan kondisi iklim di Kota Yogyakarta. *Indonesian Journal of Health Ecology*, 14(2), 124-135.

Rasmanto, M. F. (2017). Model prediksi kejadian demam berdarah dengué (DBD) berdasarkan unsur iklim di Kota Kendari tahun 2000-2015. *Jurnal Ilmiah Mahasiswa Kesehatan Masyarakat*, 1(3).

Respati, T., Raksanegara, A., Djuhaeni, H., Sofyan, A., Agustian, D., Faridah, L., & Sukandar, H. (2017). Berbagai faktor yang memengaruhi kejadian demam berdarah dengué di Kota Bandung. *ASPIRATOR-Journal of Vector-borne Disease Studies*, 9(2), 91-96.

Satrio, C. B. A. D. W. N. B. U., & H. N. Satrio. (2021). Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET. *Procedia Computer Science*, 179, 524-532.

Setiawan, H. U. E., & Setiawan, A. F. H. (2020). Penerapan Model Forecasting Arima Dan Artificial Neural Network Di Kabupaten Sragen. *Majalah Ilmiah Bahari Jogja*, 18(2), 64-78.

Team, R. (2009). Penyebaran Penyakit Demam Berdarah Berdasarkan Faktor Cuaca dan Iklim di Wilayah Kerja Puskesmas Ciwaruga Kabupaten Bandung Barat. *Jurnal Keperawatan Unpad*, 11(20).