



Artikel Penelitian

**Article history:**

Received 17 January 2024

Revised 29 January 2024

Accepted 13 February 2024

**Kata Kunci:**Karbon monoksida;  
Time series;  
Prediksi**Keywords:**Carbon monoxide;  
Time series;  
Prediction**INDEXED IN**SINTA - Science and  
Technology Index  
Crossref  
Google Scholar  
Garba Rujukan Digital: Garuda**CORRESPONDING  
AUTHOR**Ni Made Orcidia Wulaning Sari  
Prodi Meteorologi, Sekolah Tinggi  
Meteorologi Klimatologi dan Geofisika,  
Indonesia**EMAIL**[wulanningsari@gmail.com](mailto:wulanningsari@gmail.com)**OPEN ACCESS**

E ISSN 2623 - 2022

## Prediksi Karbon Monoksida Menggunakan Model *Machine Learning* Berdasarkan Perbandingan Model *Time Series* Studi Kasus DKI Jakarta

### *Carbon Monoxide Prediction Using Machine Learning Model Based on Time Series Model Comparison DKI Jakarta Case Study*

Ni Made Orcidia Wulaning Sari<sup>1\*</sup>, Hani Elindra<sup>2</sup>, Agung Hari Saputra<sup>3</sup><sup>1</sup> Program Studi Meteorologi, STMKG| [wulanningsari@gmail.com](mailto:wulanningsari@gmail.com)<sup>2</sup> Program Studi Meteorologi, STMKG| [hanielindra.ha@gmail.com](mailto:hanielindra.ha@gmail.com)<sup>3</sup> Program Studi Meteorologi, STMKG| [agung.hs@stmkg.ac.id](mailto:agung.hs@stmkg.ac.id)

**Abstrak:** DKI Jakarta sebagai pusat kegiatan ekonomi, merupakan daerah dengan dinamika aktivitas dan tingkat kepadatan penduduk yang sangat tinggi, yang mana berpengaruh terhadap kualitas udara yang ada di wilayah tersebut. Rata - rata partikel PM 2.5 di Jakarta sebesar 160, level yang termasuk berbahaya bagi kesehatan manusia. Penelitian ini bertujuan untuk mengetahui dan menguji coba tingkat keakuratan model prediksi *time series* yang didapatkan dari proses perbandingan model pada *library Pycaret* yang dilakukan pada dalam memprediksi parameter pencemar udara berupa gas karbon monoksida di wilayah DKI Jakarta. Data yang digunakan pada penelitian ini adalah data ISPU DKI Jakarta selama 11 tahun dari 2010 hingga 2021. Hasil penelitian menunjukkan bahwa perbandingan dari model *Machine Learning* dimana didapatkan 3 model terbaik yaitu model *Huber Regressor*, model *Linear Regressor* dan *Ridge Regressor* untuk memprediksi karbon monoksida. Dari hasil keseluruhan, model *Huber Regressor* masih belum dapat menghasilkan prediksi yang akurat dan optimal. Terbukti dari nilai MAE sebesar 5.3187, nilai RMSE sebesar 8.9838, nilai MASE sebesar 0.5699, nilai RMSSE sebesar 0.6571, serta sebesar 0.2364 dan 0.2061 untuk MAPE dan SMAPE. Model masih memiliki banyak keterbatasan, terutama dalam menangani *outlier* atau peristiwa tak terduga sehingga model dan metode pada penelitian ini masih membutuhkan penyesuaian serta pengembangan lebih lanjut.

**Abstract:** DKI Jakarta as the center of economic activity, is an area with very high activity dynamics and population density, which affects the air quality in the region. The average PM 2.5 particle in Jakarta is 160, a level that is considered dangerous for human health. This study aims to determine and test the level of accuracy of time series prediction models obtained from the process of comparing models carried out on *Pycaret* library in predicting air pollutant parameters in the form of carbon monoxide gas in the DKI Jakarta area. The data used in this study is DKI Jakarta's ISPU data for 11 years from 2010 to 2021. The results showed that the comparison of *Machine Learning* models where the 3 best models were obtained, namely the *Huber Regressor* model, the *Linear Regressor* model and the *Ridge Regressor* to predict carbon monoxide. From the overall results, the *Huber Regressor* model is still unable to produce accurate and optimal predictions. It is evident from the MAE value of 5.3187, RMSE value of 8.9838, MASE value of 0.5699, RMSSE value of 0.6571, as well as 0.2364 and 0.2061 for MAPE and SMAPE. The model still has many limitations, especially in dealing with outliers or unexpected events so that the models and methods in this study still require further adjustment and development.

**Jurnal Kolaboratif Sains (JKS)**

Doi: 10.56338/jks.v7i3.4819

Pages: 1116 - 1128

## LATAR BELAKANG

Hubungan antara alam dengan biota yang hidup di dalamnya merupakan kesatuan yang saling timbal balik satu sama lain. Pola kehidupan dan aktivitas manusia dipengaruhi oleh alam sebagai faktor utamanya. Sebaliknya, aktivitas yang dilakukan oleh manusia juga mempengaruhi kondisi alam (Akuba et al., 2022). Pada era modern saat ini, pertumbuhan ekonomi dan urbanisasi yang cukup tinggi di wilayah perkotaan maupun sub perkotaan menyebabkan meningkatnya penggunaan bahan bakar seperti pada pembangkit listrik, tungku - tungku industri, dan transportasi. Pembakaran bahan bakar ini merupakan sumber polutan yang mencemari udara, salah satu gas polutan yang dominan adalah karbon monoksida (CO).

DKI Jakarta yang merupakan sentral kegiatan ekonomi dan industri nasional, tercatat oleh Badan Pusat Statistik sebagai provinsi dengan aktivitas penduduk yang sangat padat. Tingginya tingkat kepadatan penduduk dan dinamika aktivitas yang ada di Jakarta dapat memberikan pengaruh terhadap kualitas udara yang ada di wilayah tersebut (Suryoprayogo et al., 2022) (Indriyaningtyas et al., 2021). Urbanisasi dan industrialisasi pesat yang terjadi di Jakarta menyebabkan masalah polusi udara (Rafferti et al., 2023). Berdasarkan data Indeks Kualitas Udara (IKU), Jakarta mencatat rata-rata partikel PM 2.5 sebesar 160, level yang dianggap berbahaya bagi kesehatan manusia. Salah satu faktor yang memperburuk kondisi udara di kota ini adalah jumlah kendaraan yang tinggi karena gas sisa pembuangan yang dihasilkan dari asap kendaraan bermotor akan menimbulkan polusi berupa gas - gas polutan, salah satunya adalah karbon monoksida (CO) yang mencemari udara (Suryoprayogo et al., 2022).

Karbon monoksida (CO) merupakan sebuah gas dengan ciri yang tidak berwarna, tidak berbau, mudah terbakar dan sangat beracun. karbon monoksida dihasilkan dari proses pembakaran material seperti batu bara, kayu, bahan bakar minyak, dan zat - zat organik lainnya yang berbahan dasar karbon (Maryanto et al., 2009). Dampak negatif dari karbon monoksida terhadap kesehatan manusia yaitu karbon monoksida dapat menempel pada haemoglobin darah dan membentuk karboksi hemoglobin (COHb). Semakin tinggi kadar CO yang dihirup oleh paru - paru maka makin fatal akibatnya, bahkan hingga menyebabkan kematian (Budiyono, 2001). Efek yang dirasakan dengan meningkatnya kadar gas CO di dalam darah antara lain yaitu mual, pusing kepala, sesak nafas, gangguan penglihatan, menurunnya konsentrasi, hilang kesadaran, dan jika terus berlanjut akan menyebabkan kematian. Paparan dalam menahun dapat berakibat pada timbulnya gejala gangguan saraf, infark otak, infark jantung dan kematian bayi dalam kandungan (Maryanto et al., 2009).

Melakukan prediksi dapat menjadi hal yang bermanfaat sebagai upaya untuk memantau dan menangani isu kualitas udara kedepannya, sehingga langkah preventif dapat diambil. *Artificial intelligence* merupakan sebuah terobosan modern yang saat ini telah banyak dipakai untuk membuat prediksi, dengan metode yang paling populer adalah *Machine Learning* (Akhtar & Moridpour, 2021). Keunggulan dari *Machine Learning* anatara lain yaitu memberikan implementasi yang lebih mudah dengan biaya komputasi yang rendah, serta pelatihan, validasi, pengujian, dan evaluasi yang cepat, dengan kinerja yang tinggi dibandingkan dengan model fisik, dan relatif lebih tidak rumit (Mosavi et al., 2018). Semakin baik algoritmanya maka prediksi akan semakin akurat. Pemodelan tipe *time series* merupakan salah satu model yang banyak digunakan untuk membuat prediksi. Model *time series* mengamati perilaku observasi yang diambil dari waktu ke waktu secara berurutan (Hadiansyah, 2017). Pada penelitian - penelitian sebelumnya, metode *time series* telah banyak diimplementasikan dalam membuat prediksi yang berkaitan dengan waktu, seperti yang dilakukan oleh Syaifulloh (2021) dalam memprediksi kadar karbon monoksida di Surabaya dan Paul (2023) yang membahas prediksi *time series* nitrogen dioksida.

Pada penelitian kali ini dilakukan perbandingan *time series* dengan tujuan untuk mencari model terbaik yang kemudian akan digunakan sebagai model prediksi atau forecast. Hasil visualisasi dan *time series* yang didapatkan kemudian akan dianalisis dan dievaluasi untuk mengetahui akurasi dari model yang dipilih. Penelitian ini bertujuan untuk mengetahui dan menguji coba tingkat keakuratan dari model prediksi yang diperoleh dari proses perbandingan *time series* pada sistem

*Machine Learning* dalam memprediksi parameter pencemar udara berupa gas karbon monoksida di wilayah DKI Jakarta.

## METODE

### Teknik Pengumpulan Data

Penelitian terkait prediksi yang dilakukan pada Januari 2022 mengambil sampel lokasi pada provinsi DKI Jakarta yang terletak di  $6^{\circ}12'$  lintang selatan dan  $106^{\circ}48'$  bujur timur. DKI Jakarta memiliki luas wilayah daratan  $664,01 \text{ km}^2$  dan luas wilayah lautan  $6977,5 \text{ km}^2$  (Puji et al., 2022). Luas Daerah ini mencakup kota administrasi Jakarta Pusat, Jakarta Utara, Jakarta Barat, Jakarta Timur, dan Jakarta Selatan. Pengambilan data untuk penelitian diawali dengan mengakses *Website* Open Data Jakarta (<https://data.jakarta.go.id>), di mana terdapat data Indeks Standar Pencemaran Kualitas Udara (ISPU) DKI Jakarta dari tahun 2010 sampai 2021 yang dapat diakses dan diunduh secara gratis. Dataset tersebut berisi mengenai Indeks Standar Pencemaran Udara (ISPU) yang diukur dari lima stasiun pemantau kualitas udara (SPKU) yang ada di Provinsi DKI Jakarta.

### Pengolahan Data

#### a. Pre-processing Data

Setelah mengunduh data ISPU DKI Jakarta 2010 - 2021, data kemudian akan dipindahkan dan digabungkan dalam satu file *Microsoft Excel*. File Excel ini kemudian akan disimpan pada *Google Drive* lalu dijalankan di platform *Google Collaboratory* untuk diproses lebih lanjut.

#### b. Processing Data menggunakan *Machine Learning*

Proses pembuatan prediksi atau forecast dilakukan menggunakan metode *Machine Learning*, dengan platform *Google Colaboratory* dan library *Pycaret* serta library *pandas* dan *matplotlib*. *Pycaret* merupakan library lowcode untuk ML dalam Python (Anwar, et al, 2021). *Lowcode* berarti hanya membutuhkan sedikit kode dalam penggunaannya (Anwar et al., 2021). *Pycaret* pada dasarnya merupakan wrapper yang memanfaatkan beberapa library dan framework ML seperti *Scikit-learn* (Bengfort & Bilbro, 2019), *XGBoost* (Azan et al., 2023), *Microsoft LightGBM* (Zhang, 2023), *SpaCy* (Schmitt et al., 2019) dan lain-lain. Cakupan proses dan fungsi yang dapat dilakukan oleh *Pycaret* antara lain adalah penyiapan data, pemodelan, perbandingan model, parameter tuning, dan abstraksi model (Anwar, et al, 2021). Dari sekian banyaknya fungsi pada *Pycaret*, ada kalanya fungsi yang digunakan hanya sebagian. Pada penelitian ini, fungsi utama yang digunakan dari *Pycaret* adalah untuk memilih model terbaik dari sekian banyak model berdasarkan nilai akurasi.

#### c. Pengaturan *Hyperparameter*

Setelah *processing* data selesai dilakukan, selanjutnya dilakukan pengaturan *Hyperparameter* yang bertujuan untuk memastikan model memiliki hasil yang optimal dan sesuai dengan target. *Hyperparameter* yang diatur antara lain yaitu *target*, *user\_defined\_seasonal\_period(s)*, *session\_id*, *fold*, *seasonality\_present*, *remove\_harmonics*, *seasonal\_period*, *fold\_strategy*, *max\_sp\_to\_consider*, *harmonic\_order\_method*, dan *num\_sps\_to\_use*. Berikut merupakan tabel *Hyperparameter* beserta konfigurasinya

**Tabel 1.** Konfigurasi *Hyperparameter*

| <b>Hyperparameter</b>                  | <b>Konfigurasi</b>       |
|--|--------------------------|
| <i>target</i>                          | co                       |
| <i>user_defined_seasonal_period(s)</i> | 7                        |
| <i>session_id</i>                      | 123                      |
| <i>fold</i>                            | 5                        |
| <i>seasonality_present</i>             | True                     |
| <i>remove_harmonics</i>                | True                     |
| <i>seasonal_period_</i>                | 7                        |
| <i>fold_strategy</i>                   | <i>sliding</i>           |
| <i>max_sp_to_consider</i>              | 48                       |
| <i>harmonic_order_method</i>           | <i>harmonic_strength</i> |
| <i>num_sps_to_use</i>                  | 1                        |

#### d. Jenis - Jenis Model

Berikut merupakan jenis- jenis model yang digunakan untuk perbandingan model pada penelitian kali ini.

**Tabel 2.** Daftar Model Prediksi *Machine Learning*

| <b>Nomor</b> | <b>Model</b>           | <b>Referensi</b>               |
|--------------|------------------------|--------------------------------|
| 1            | <i>Huber</i>           | Cavazza & Murino, (n.d)        |
| 2            | <i>Linear</i>          | Dette & Munk (1998)            |
| 3            | <i>Ridge</i>           | Vlaming & Groenen (2015)       |
| 4            | <i>Bayesian Ridge</i>  | Efendi & Effrihan (2017)       |
| 5            | <i>Elastic Net</i>     | Hans (2011)                    |
| 6            | <i>Lasso</i>           | Uniejewski & Weron (2018)      |
| 7            | <i>Lasso Regressor</i> | Ranstam & Cook (2018)          |
| 8            | <i>Auto ARIMA</i>      | Yermal & Balasubramanian(2017) |

#### e. Visualisasi dan Analisis Grafik

Prediksi yang dibuat oleh model akan menghasilkan produk visual berupa beberapa set grafik *time series*. Pada set grafik *time series* terdapat grafik dekomposisi yang terdiri dari *trend*, *seasonal*, dan *residual*, serta grafik *diagnostic*, grafik *difference*, dan grafik *actual vs forecast*. Setelah itu, analisis dan interpretasi secara kualitatif akan dilakukan terhadap hasil visualisasi grafik yang telah didapatkan.

#### f. Evaluasi Model

Setelah dilakukan analisis dari grafik *time series*, langkah selanjutnya adalah melakukan evaluasi terhadap nilai error yang ada pada model untuk memastikan bahwa hasil yang diberikan model optimal dan akurat. Evaluasi dilakukan dengan melihat nilai metrik *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), MAPE (*Mean Absolute Percentage Error*) dan SMAPE (*Symmetric Mean Absolute Percentage Error*), dan MASE (*Mean Absolute Scaled Error*) yang menunjukkan tingkat kesalahan yang ada pada prediksi model. Secara matematis, nilai - nilai tersebut dapat dituliskan sebagai berikut :

$$RMSE = \sqrt{\frac{1}{T} \sum_{i=1}^T (y_i - \hat{y}_i)^2} \quad (1)$$

$$MAE = \frac{1}{T} \sum_{i=1}^T |y_i - \hat{y}_i| \quad (2)$$

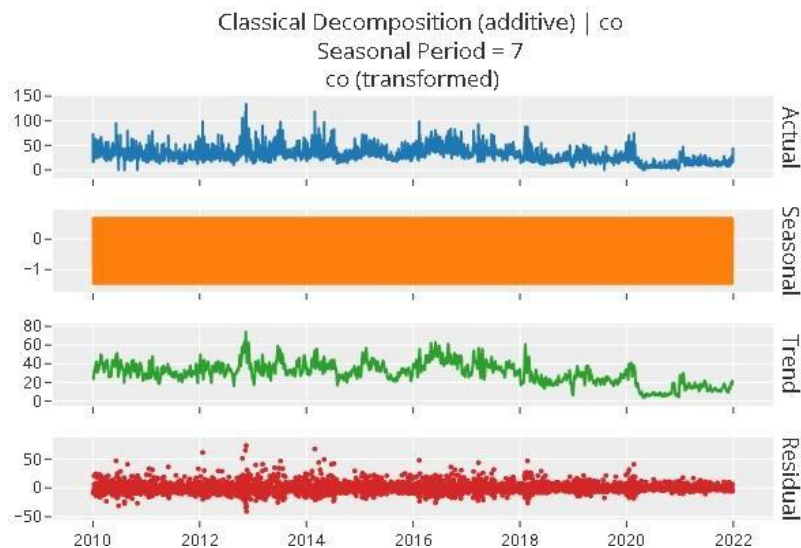
$$MAPE = 100\% \frac{1}{T} \sum_{i=1}^T \frac{|y_i - \hat{y}_i|}{|y_i|} \quad (3)$$

$$SMAPE = \frac{1}{T} \sum_{i=1}^T \frac{|y_i - \hat{y}_i|}{(\frac{1}{2}(|y_i| + |\hat{y}_i|))} \quad (4)$$

$$MASE = \frac{e_i}{\frac{1}{T-1} \sum_{i=2}^T |y_i + \hat{y}_{i-1}|} \quad (5)$$

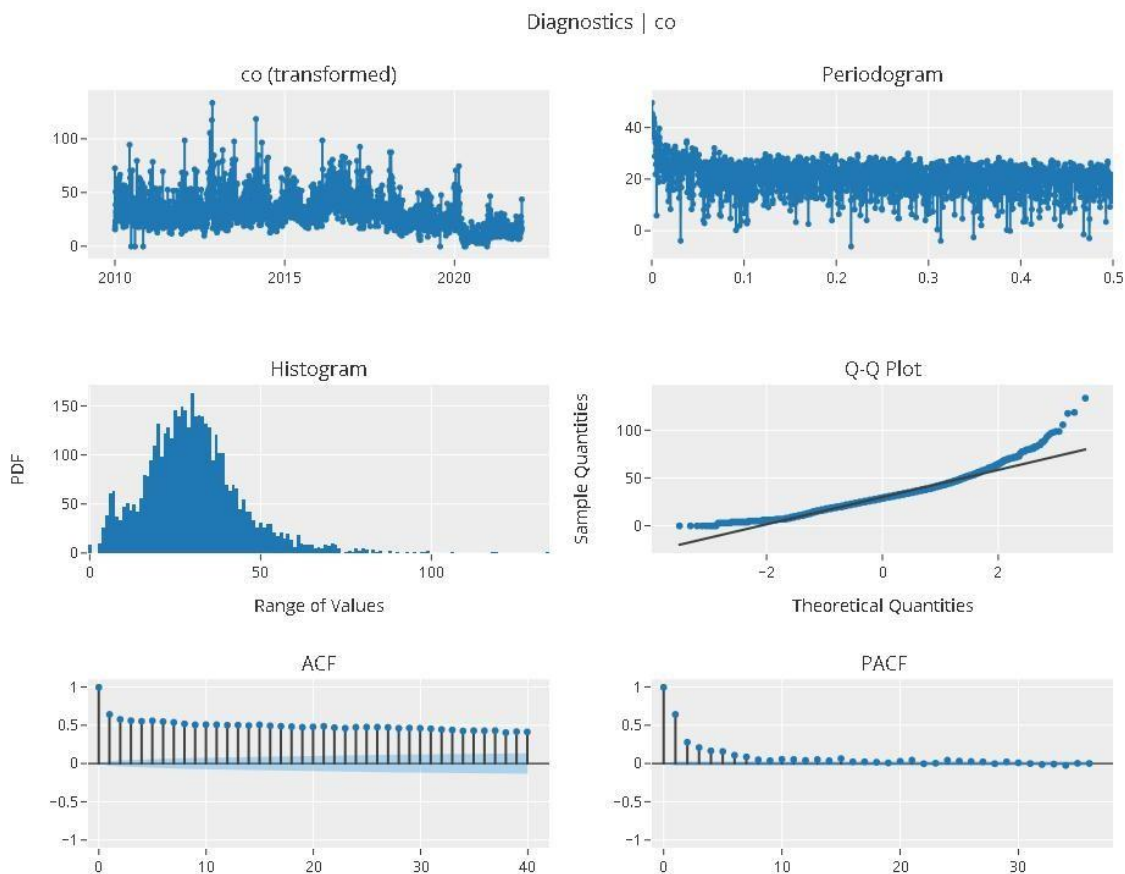
di mana  $t$  adalah ukuran data,  $y$  prediksi saat  $i$ ,  $y_i$  data saat  $i$ , dan adalah kesalahan antara nilai yang diramalkan dan nilai aktual. RMSE adalah ukuran yang sering digunakan untuk mengukur perbedaan antara nilai yang diprediksi oleh model dan data yang diamati. Model ini adalah akar kuadrat dari rata-rata kesalahan kuadrat. MAE adalah rata-rata dari kesalahan absolut antara prediksi dan nilai sebenarnya, yang memberikan rata-rata kesalahan peramalan absolut. MAPE mengukur akurasi sebagai persentase yang dapat dihitung sebagai persen kesalahan absolut kumulatif untuk setiap kerangka waktu, sebagai nilai aktual dikurangi nilai prediksi dibagi dengan nilai aktual. Dengan kata lain, ini menggambarkan kesalahan rata-rata dalam bentuk persentase (Luo et al., 2021). SMAPE adalah versi modifikasi dari MAPE di mana pembagiannya adalah setengah dari jumlah nilai aktual dan prakiraan. Ukuran lain, MASE diperoleh dengan menskalakan kesalahan prediksi berdasarkan kesalahan absolut rata-rata dalam sampel dan dapat mengatasi kelemahan MAPE yang terkadang menghasilkan nilai yang tidak terbatas atau tidak terdefinisi (Kim S & Kim H, 2016).

## HASIL DAN PEMBAHASAN



Gambar 1. Dekomposisi klasik (*additive*) time series CO Jakarta 2010-2021

Gambar 1 merupakan dekomposisi klasik (additive) sebuah *time series*. Metode dekomposisi umumnya memisahkan rangkaian *time series* dengan mengambil fitur atau pola yang mendasarinya (Iwana & Uchida, 2021). Pada grafik, *time series* menjadi beberapa komponen dasar, yaitu: tren, seasonalitas, dan residu. Panel atas (*actual*) menunjukkan data asli yang telah diubah (*transformed*) dari *time series* CO. Data terlihat fluktuatif dengan beberapa puncak dan lembah yang menonjol yang mungkin mengindikasikan pola musiman atau pengaruh eksternal lainnya. Panel kedua (*seasonal*) menampilkan komponen musiman dari data, dengan periode musiman disetel ke 7. Terlihat bahwa ada pola yang konsisten dan berulang yang bisa jadi berkaitan dengan hari dalam seminggu atau siklus mingguan lainnya, mengingat nilai periode musiman yang digunakan. *Trend* pada panel ketiga menggambarkan tren atau arah jangka panjang dari data. Terlihat bahwa ada periode kenaikan dan penurunan yang menunjukkan adanya perubahan bertahap yang mungkin terkait dengan faktor-faktor eksternal atau internal terhadap sistem yang dianalisis. Paling bawah menunjukkan *residual*, yaitu bagian dari data yang tidak dijelaskan oleh komponen musiman atau *trend*. Dalam grafik ini, residu terlihat cukup acak dengan beberapa *outlier* yang menonjol. Residu yang berdekatan dengan garis nol dan terdistribusi secara acak menunjukkan bahwa model dekomposisi telah berhasil menangkap sebagian besar informasi dari data asli. Secara keseluruhan, dekomposisi ini menunjukkan bahwa ada pola musiman yang kuat dan trend yang jelas dalam data. Residu yang relatif kecil dan acak menunjukkan bahwa model dekomposisi adalah cukup efektif dalam menangkap pola dalam data.

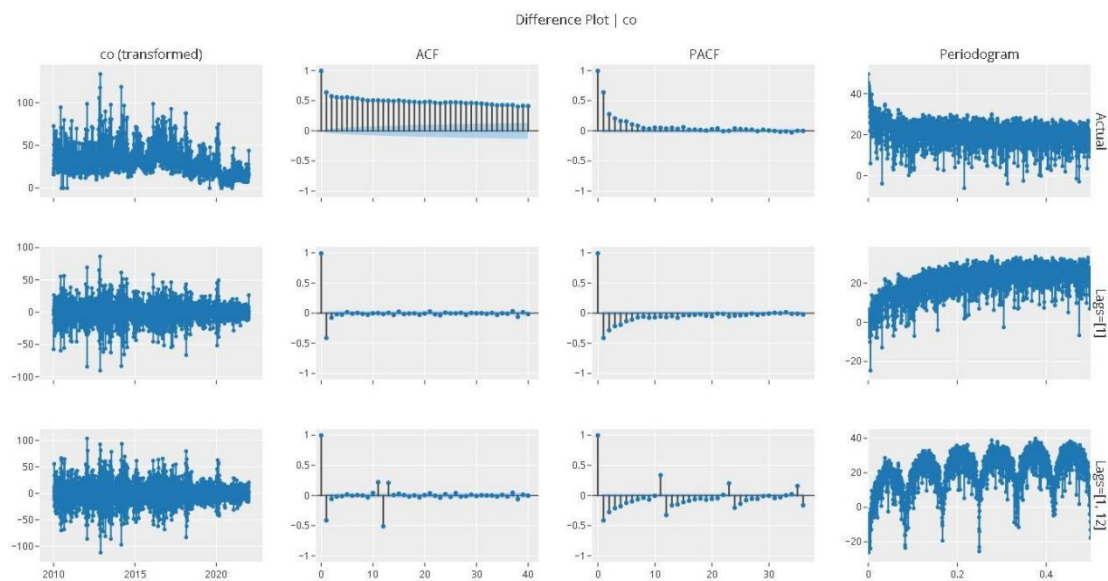


**Gambar 2.** Diagnostik CO Jakarta 2010-2021: Plot *time series* (atas kiri), Periodogram (atas kanan), Histogram (tengah kiri), Plot Q-Q (Tengah kanan), ACF (*Autocorrelation Function*) (bawah kiri), dan PACF (*Partial Autocorrelation Function*) (bawah kanan)

Gambar 2 menampilkan serangkaian plot diagnostik untuk data *time series* konsentrasi CO, dari plot seri waktu, terlihat bahwa data memiliki variasi yang signifikan dengan beberapa puncak yang

menonjol, menandakan kemungkinan adanya insiden sesekali yang menghasilkan nilai-nilai tinggi. Meski tidak ada pola tren yang jelas terlihat, fluktuasi ini mungkin mengandung informasi penting tentang perilaku dinamis dari konsentrasi CO yang diukur. Periodogram tidak menunjukkan adanya puncak menonjol, yang mengindikasikan tidak adanya frekuensi dominan atau siklus musiman yang konsisten dalam data. Hal ini bisa mengimplikasikan bahwa konsentrasi CO tidak dipengaruhi oleh pola musiman yang jelas atau siklus yang berulang dengan interval waktu tetap. Histogram dari data mengungkapkan distribusi yang miring ke kanan dengan ekor yang panjang, menandakan bahwa sebagian besar pengamatan memiliki nilai yang rendah, ada beberapa kejadian yang sangat tinggi yang dapat dianggap sebagai *outlier*. Ini menunjukkan bahwa konsentrasi CO cenderung rendah secara umum tetapi dapat terjadi lonjakan tajam. Q-Q plot menunjukkan bahwa data tidak sepenuhnya mengikuti distribusi normal, terutama pada nilai-nilai ekstrem yang bisa disebabkan oleh kemunculan *outlier* yang telah diidentifikasi dalam histogram. Plot ini menunjukkan bahwa model statistik yang mengasumsikan normalitas mungkin tidak cocok dengan sempurna untuk data ini.

Dalam analisis *autocorrelation* dengan ACF, terdapat penurunan cepat dalam korelasi, memperlihatkan bahwa korelasi antara nilai-nilai *time series* sebagian besar terbatas pada interval waktu yang sangat dekat dengan korelasi yang menurun dengan cepat, menunjukkan bahwa informasi masa lalu tidak terlalu mempengaruhi nilai-nilai yang akan datang. PACF mendukung hal ini dengan menunjukkan tidak adanya korelasi parsial yang signifikan pada lag yang lebih tinggi, yang mengarahkan bahwa model prediktif yang kompleks tidak diperlukan.



**Gambar 3.** Perbedaan plot *time series* CO, ACF, PACF, dan periodogram pada data yang sebenarnya, lags=[1], dan lags[1,12]

Gambar 3 menampilkan analisis *time series* untuk konsentrasi CO (*transformed*) yang disajikan dalam tiga baris yang berbeda, masing-masing dengan interpretasi yang berbeda berdasarkan lag yang diaplikasikan pada data. Di baris pertama, menampilkan plot dari data aktual *time series* CO, plot ACF, PACF, dan periodogram. Baris kedua menampilkan data yang telah di-*differencing* dengan lag 1, yang merupakan upaya untuk mengurangi non-stasioneritas dengan menghilangkan tren dan ketergantungan

langsung pada waktu. Plot ACF dan PACF pada data ini menunjukkan penurunan korelasi dari aktual, yang menandakan bahwa diferensiasi telah berhasil mengurangi *autocorrelation* dan membantu dalam mengidentifikasi struktur yang mungkin lebih stasioner dalam data. Periodogram menunjukkan distribusi frekuensi yang seragam, mengindikasikan pengurangan kemungkinan musimanitas atau tren dalam data. Pada baris ketiga, di mana data telah di-*differencing* dengan lag 1 dan tambahan lag 12 dilakukan untuk menangani musimanitas yang mungkin tidak teridentifikasi pada diferensiasi lag 1 saja. ACF menunjukkan pola korelasi yang berubah signifikan, dengan beberapa korelasi yang tetap pada lag tertentu, yang mungkin mencerminkan sisa musimanitas atau pola yang belum dihilangkan. PACF menunjukkan beberapa korelasi parsial yang masih signifikan, yang bisa menjadi informasi berharga untuk pemodelan lanjutan. Periodogram yang berubah juga menandakan adanya perubahan dalam karakteristik frekuensi setelah diferensiasi lebih lanjut.

Tabel 3. Perbandingan Model

| Model                              | MASE   | MAE     | RMSE    | MAPE   | SMAPE  |
|------------------------------------|--------|---------|---------|--------|--------|
| <i>Huber</i>                       | 0.4198 | 3.9253  | 4.8890  | 0.3066 | 0.2820 |
| <i>Linear</i>                      | 0.4461 | 4.1714  | 5.0867  | 0.3756 | 0.2938 |
| <i>Ridge</i>                       | 0.4461 | 4.1714  | 5.0867  | 0.3756 | 0.2938 |
| <i>Bayesian Ridge</i>              | 0.4465 | 4.1744  | 5.0892  | 0.3761 | 0.2940 |
| <i>Elastic Net</i>                 | 0.4479 | 4.1881  | 5.0998  | 0.3782 | 0.2949 |
| <i>Lasso</i>                       | 0.4496 | 4.2035  | 5.1112  | 0.3805 | 0.2959 |
| <i>Lasso Regressor</i>             | 0.4496 | 4.2035  | 5.1113  | 0.3805 | 0.2959 |
| <i>Auto ARIMA</i>                  | 0.4523 | 4.2293  | 5.0405  | 0.3314 | 0.3060 |
| <i>Theta Forecaster</i>            | 0.4527 | 4.2329  | 5.1194  | 0.3354 | 0.3043 |
| <i>ETS</i>                         | 0.4575 | 4.2788  | 5.2406  | 0.3342 | 0.3061 |
| <i>Exponential Smoothing</i>       | 0.4637 | 4.3360  | 5.2763  | 0.3462 | 0.3094 |
| <i>Croston</i>                     | 0.4689 | 4.3843  | 5.2019  | 0.3470 | 0.3164 |
| <i>STLF</i>                        | 0.4948 | 4.6256  | 5.8829  | 0.3778 | 0.3327 |
| <i>Gradient Boosting</i>           | 0.5039 | 4.7115  | 5.7193  | 0.4447 | 0.3206 |
| <i>Extreme Gradient Boosting</i>   | 0.5145 | 4.8119  | 5.7374  | 0.4219 | 0.3195 |
| <i>Naive Forecaster</i>            | 0.5150 | 4.8167  | 5.6708  | 0.3822 | 0.3480 |
| <i>Light Gradient Boosting</i>     | 0.5415 | 5.063   | 6.0832  | 0.4817 | 0.3354 |
| <i>Random Forest</i>               | 0.5470 | 5.1136  | 6.2057  | 0.4758 | 0.3382 |
| <i>Seasonal Naive Forecaster</i>   | 0.5742 | 5.3667  | 6.7444  | 0.4444 | 0.3886 |
| <i>ARIMA</i>                       | 0.5756 | 5.3794  | 6.7637  | 0.4471 | 0.3936 |
| <i>K Neighbors</i>                 | 0.5828 | 5.4491  | 6.7861  | 0.5136 | 0.3509 |
| <i>Extra Trees</i>                 | 0.5894 | 5.5121  | 6.3806  | 0.4979 | 0.3589 |
| <i>Decision Tree</i>               | 0.6019 | 5.6315  | 7.9733  | 0.4610 | 0.3458 |
| <i>Orthogonal Matching Pursuit</i> | 0.6210 | 5.8053  | 6.6801  | 0.5562 | 0.3813 |
| <i>Polynomial Trend Forecaster</i> | 0.6661 | 6.2270  | 7.0888  | 0.5971 | 0.4028 |
| <i>Grand Means Forecaster</i>      | 1.7207 | 16.0877 | 16.646  | 1.4085 | 0.7537 |
| <i>AdaBoost</i>                    | 2.0293 | 18.9705 | 20.5099 | 1.6029 | 0.7934 |

Pada tabel 3, MASE (*Mean Absolute Scaled Error*) merupakan ukuran dengan nilai yang lebih rendah menunjukkan performa yang lebih baik. Pada metrik ini, model seperti *Huber Regressor* dan *Theta Forecaster* menunjukkan hasil yang menjanjikan, mengindikasikan bahwa prediksi sejalan dengan nilai aktual sambil memperhitungkan skala. MAE (*Mean Absolute Error*) dan RMSE (*Root*



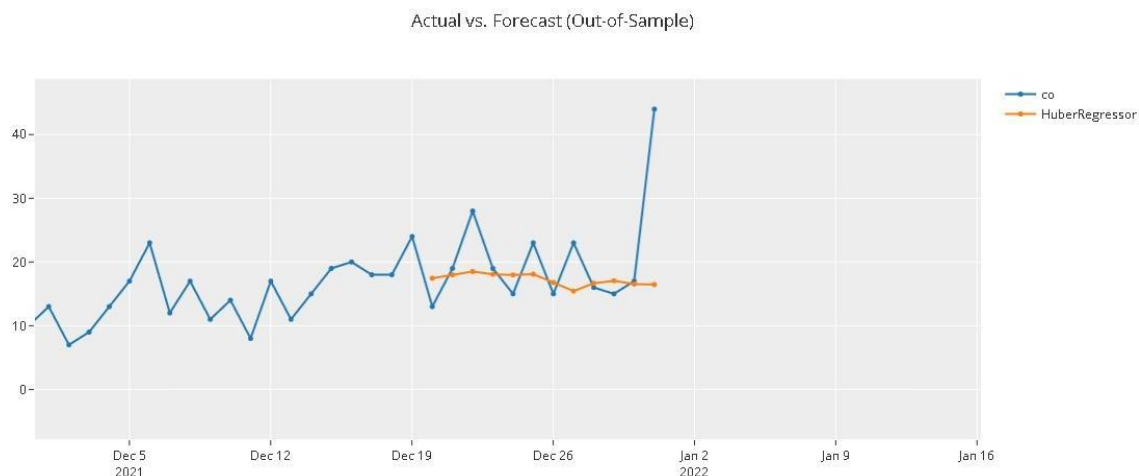
*Mean Squared Error*) memberikan ukuran absolut dari akurasi prediksi, dengan nilai yang lebih rendah menandakan prediksi yang lebih akurat. Di sini, *Huber Regressor* sekali lagi menunjukkan performa yang baik, bersama dengan model *Linear* dan *Ridge*, yang menunjukkan kapasitas yang kuat dalam menangkap kecenderungan sentral data tanpa terlalu dipengaruhi oleh nilai-nilai ekstrim. *MAPE* (*Mean Absolute Percentage Error*) dan *SMAPE* (*Symmetric Mean Absolute Percentage Error*) menawarkan perspektif berbasis persentase dari kesalahan, yang berguna untuk membandingkan performa yang berbeda. Model dengan nilai lebih rendah pada metrik ini, seperti *Huber Regressor* dan *Exponential Smoothing*, akan menguntungkan dalam skenario di mana akurasi proporsional sangat penting.

**Tabel 4.** Evaluasi Model Terhadap *Data Training*

|             | <b>cutoff</b> | <b>MASE</b> | <b>MAE</b> | <b>RMSE</b> | <b>MAPE</b> | <b>SMAPE</b> |
|-------------|---------------|-------------|------------|-------------|-------------|--------------|
| 0           | 20/10/2021    | 0.2540      | 2.3858     | 3.1494      | 0.1390      | 0.1377       |
| 1           | 01/11/2021    | 0.7647      | 7.1665     | 8.8195      | 0.5986      | 0.4039       |
| 2           | 13/11/2021    | 0.8401      | 7.8485     | 9.1040      | 0.9364      | 0.5610       |
| 3           | 25/11/2021    | 0.4968      | 4.6351     | 5.0931      | 0.4645      | 0.3614       |
| 4           | 07/12/2021    | 0.3518      | 3.2790     | 4.2499      | 0.2700      | 0.2132       |
| <b>Mean</b> |               | 0.5415      | 5.063      | 6.0832      | 0.4817      | 0.3354       |
| <b>SD</b>   |               | 0.2279      | 2.1315     | 2.4315      | 0.2767      | 0.1485       |

Berdasarkan tabel 4, model menunjukkan nilai evaluasi model pada *data training*. Dengan *Mean Absolute Scaled Error* (MASE) rata-rata sekitar 0.4198, hal ini mengindikasikan bahwa model memiliki kemampuan prediksi yang baik untuk sebagian besar waktu, meskipun nilai ini juga menunjukkan adanya ruang untuk peningkatan karena nilai MASE idealnya mendekati 0. Ketika mempertimbangkan *Mean Absolute Error* (MAE) dan *Root Mean Squared Error* (RMSE) dengan nilai rata-rata masing-masing 3.9253 dan 4.8890, terlihat bahwa model memiliki tingkat kesalahan absolut dan kesalahan kuadrat yang signifikan dalam beberapa prediksinya. Hal ini menunjukkan bahwa model tersebut mengalami kesulitan dalam menangani kompleksitas tertentu dalam data sehingga menghasilkan kesalahan yang lebih besar dalam beberapa prediksi yang membuat prediksi menjadi kurang akurat. Pada model *Mean Absolute Percentage Error* (MAPE) dan *Symmetric Mean Absolute Percentage Error* (SMAPE) dengan nilai rata-rata 0.3066 dan 0.2820 menunjukkan bahwa model memiliki kesalahan yang cukup tinggi terhadap nilai sebenarnya yang diprediksi. Nilai ini menunjukkan bahwa terdapat variasi prediksi yang cukup luas dari nilai yang sebenarnya.

*Standar Deviasi* (SD) menunjukkan variabilitas kesalahan model. Semakin rendah nilai ini, semakin konsisten model dalam melakukan prediksi. Secara konsistensi, model menunjukkan variabilitas dalam performa, seperti yang ditunjukkan oleh standar deviasi dari masing-masing model. Ini menunjukkan bahwa meskipun model memiliki tingkat kesalahan yang dapat diterima secara umum, performanya bervariasi di berbagai kasus atau periode waktu. Hal ini bisa jadi karena model sensitif terhadap kondisi tertentu atau karena fluktuasi intrinsik dalam data yang diprediksi.



**Gambar 4.** Perbandingan nilai aktual dan prediksi yang dihasilkan oleh model *Huber Regressor*

Gambar 4 menunjukkan sebuah perbandingan antara nilai-nilai aktual yang diwakili oleh garis biru dan nilai-nilai yang diprediksi oleh *Huber Regressor* yang dipresentasikan melalui garis oranye. *Huber Regressor* adalah model yang banyak digunakan karena set data ini sering kali memiliki kemiripan dan kehalusan yang tinggi di antara data-data (Xin et al., 2022). Hal ini mungkin menjelaskan mengapa prediksi terlihat tidak terpengaruh oleh puncak ekstrem dalam data aktual dan cenderung lebih rata. Pada titik terakhir, prediksi model gagal merefleksikan lonjakan tajam yang terlihat pada data aktual, yang menandakan kemungkinan keterbatasan model dalam menangani *outlier* atau peristiwa tak terduga. Ini bisa menunjukkan bahwa model tersebut lebih baik dalam menangkap tren umum daripada menyesuaikan diri dengan perubahan mendadak atau kejadian yang tidak teratur.

**Tabel 5.** Evaluasi Model Terhadap *Data Testing*

| No | Model | Nilai  |
|----|-------|--------|
| 1  | MASE  | 0.5699 |
| 2  | MAE   | 5.3187 |
| 3  | RMSE  | 8.9838 |
| 4  | MAPE  | 0.2061 |
| 5  | SMAPE | 0.2364 |

Pada tabel 5 diperlihatkan bahwa nilai MAE, yang mengukur rata-rata kesalahan absolut dalam prediksi yang dilakukan oleh model, adalah 5.3187. Dalam konteks ini, nilai MAE menunjukkan bahwa secara rata-rata, kesalahan model dalam memprediksi adalah sebesar 5.3187 unit dari nilai sebenarnya. Nilai yang lebih rendah tersebut menunjukkan bahwa model memiliki kinerja yang lebih baik. Kemudian didapatkan nilai RMSE sebesar 8.9838 dengan nilai yang tinggi menunjukkan bahwa ada variasi yang cukup besar dalam kesalahan prediksi model. Nilai MAPE adalah 0.2061, atau 20.61% jika diubah menjadi persentase. MAPE memberikan gambaran tentang seberapa besar kesalahan prediksi relatif terhadap nilai sebenarnya. Nilai MAPE sebesar 20.61% bisa dianggap cukup tinggi, menunjukkan bahwa model mungkin tidak melakukan prediksi dengan baik dalam beberapa kasus relatif terhadap skala data. Nilai SMAPE adalah 0.2364, atau 23.64% dalam bentuk persentase. SMAPE adalah versi simetris dari MAPE dan dianggap lebih akurat ketika data memiliki nilai-nilai yang sangat rendah atau nol, karena tidak menempatkan terlalu banyak penekanan pada kesalahan persentase ketika nilai sebenarnya mendekati nol.

## KESIMPULAN

Setelah melalui proses perbandingan model melalui *Pycaret*, didapatkan bahwa model *Huber Regressor* merupakan model terbaik dari sekian banyak model lainnya, hal ini dapat dilihat dari nilai MAE (*Mean Absolute Error*) dan RMSE (*Root Mean Squared Error*), yang menunjukkan performa yang lebih baik. Meski memiliki nilai MAE dan RMSE yang rendah, Percobaan yang dilakukan terhadap model dalam membuat forecast menunjukkan bahwa Pada titik terakhir, prediksi model gagal merefleksikan lonjakan tajam yang terlihat pada data aktual, yang menandakan kemungkinan keterbatasan model dalam menangani *outlier* atau peristiwa tak terduga. Ini bisa menunjukkan bahwa model tersebut lebih baik dalam menangkap tren umum daripada menyesuaikan diri dengan perubahan mendadak atau kejadian yang tidak teratur.

## SARAN

Penelitian lebih lanjut perlu disarankan untuk melakukan analisis data yang mendalam, memastikan pemahaman yang komprehensif terhadap pola dan karakteristik data, termasuk keberadaan tren dan musimanitas. Pemilihan model harus didasarkan pada sifat data, dengan pertimbangan khusus pada model yang tahan terhadap *outlier* seperti *Huber Regressor*, atau model yang lebih canggih seperti ARIMA untuk pola musiman. Penting juga untuk menerapkan validasi silang dalam pengujian model, memastikan keandalan prediksi, serta disarankan untuk terus bereksplorasi dan terbuka terhadap metodologi baru, sambil mempertahankan pendekatan kritis terhadap hasil. Kolaborasi antar peneliti dan dokumentasi yang baik akan mendukung kemajuan ilmu peramalan dan memastikan penelitian yang berkualitas dan dapat direproduksi.

## DAFTAR PUSTAKA

- Akhtar, M., & Moridpour, S. (2021). A Review of Traffic Congestion Prediction Using Artificial Intelligence. In *Journal of Advanced Transportation* (Vol. 2021). Hindawi Limited. <https://doi.org/10.1155/2021/8878011>
- Akuba, M. N., Fernandez, S., Adrian, P., Maskim, F., Ganimel, D., & Simanullang, A. (2022). *PERBANDINGAN MODEL DALAM MEMREDIKSI KARBON MONOKSIDA DI KOTA MEDAN BULAN JANUARI 2023* (Vol. 2, Issue 2).
- Anwar, M. T., & Permana, D. R. A. (2021). Perbandingan Performa Model Data Mining untuk Prediksi Dropout Mahasiswa. *Jurnal Teknologi Dan Manajemen*, 19(2), 33–40. <https://doi.org/10.52330/jtm.v19i2.34>
- Barth, L., & Kus, D. (2020). *PRIME REPRESENTATIONS IN THE HERNANDEZ-LECLERC CATEGORY: CLASSICAL DECOMPOSITIONS*.
- Budiyono, A. (n.d.). *Pencemaran Udara : Dampak Pencemaran Udara Pada Lingkungan*.
- Cavazza, J., & Murino, V. (n.d.). *Active Regression with Adaptive Huber Loss*.
- Dette, H., & Munk, A. (1998). Validation of Linear Regression Models. *The Annals of Statistics*, 26(2), 778–800. <http://www.jstor.org/stable/120052>
- Gorr, W. (n.d.). *Estimation of Crime Seasonality: A Cross-Sectional Extension to Time Series Classical Decomposition*. <https://www.researchgate.net/publication/228823192>
- Hermawan Soekamto, T., & Perdanakusuma, D. (n.d.). *INTOKSIKASI KARBON MONOKSIDA*
- Hadiansyah, F. N. (n.d.). *Prediksi Harga Cabai dengan Pemodelan Time Series ARIMA*. <https://doi.org/10.21108/indojc.2017.21.144>
- Hans, C. (2011). Elastic Net Regression Modeling With the Orthant Normal Prior. *Journal of the American Statistical Association*. 106:496, 1383-1393, DOI: 10.1198/jasa.2011.tm09241

- Indriyaningtyas, S., Hasandy, L. R., & Dewantoro, B. E. B. (2021). DINAMIKA KONSENTRASI EMISI GAS KARBON MONOKSIDA (CO) SELAMA PERIODE PSBB MENGGUNAKAN KOMPUTASI BERBASIS CLOUD PADA GOOGLE EARTH ENGINE Studi Kasus di Provinsi DKI Jakarta, Indonesia. *MAJALAH ILMIAH GLOBE*, 23(1), 35. <https://doi.org/10.24895/mig.2021.23-1.1258>
- Iwana, B. K., & Uchida, S. (2021). An empirical survey of data augmentation for time series classification with neural networks. In *PLoS ONE* (Vol. 16, Issue 7 July). Public Library of Science. <https://doi.org/10.1371/journal.pone.0254841>
- Kanezar, A., Sasmita, M. A., & Saputra, A. H. (n.d.). *Prediksi Particulate Matter (PM 2.5) di DKI Jakarta Menggunakan XGBoost (Azan Kanezar MS dkk) Prediksi Particulate Matter (PM 2.5) di DKI Jakarta Menggunakan XGBoost*. <https://jurnal.stmkg.ac.id/index.php/jam>
- Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3), 669–679. <https://doi.org/10.1016/j.ijforecast.2015.12.003>
- Luo, J., Zhang, Z., Fu, Y., & Rao, F. (2021). Time series prediction of COVID-19 transmission in America using LSTM and XGBoost algorithms. *Results in Physics*, 27. <https://doi.org/10.1016/j.rinp.2021.104462>
- Maryanto, D., Mulasari, S. A., Dyah, S., Fakultas, K., Masyarakat, A., & Dahlan, Y. (2009). *KES MAS PENURUNAN KADAR EMISI GAS BUANG KARBON MONOKSIDA (CO) DENGAN PENAMBAHAN ARANG AKTIF PADA KENDARAAN BERMOTOR DI YOGYAKARTA*.
- Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. In *Water (Switzerland)* (Vol. 10, Issue 11). MDPI AG. <https://doi.org/10.3390/w10111536>
- Puji Lestari, Maulana Khafid Arrohman, Seny Damayanti, Zbigniew Klimont, Emissions and spatial distribution of air pollutants from anthropogenic sources in Jakarta, *Atmospheric Pollution Research*, Volume 13, Issue 9, 2022, 101521, ISSN 1309-1042.
- Rafferti Kriswandanu, A., Ammar Muhtadi, M., Suprayitno, E., & Wardita, Y. (2023). Effects of Urbanization, Transport Infrastructure, Air Quality, and Health Outcomes on the Quality of Life of Jakarta City Population Article Info ABSTRAK. In *Jurnal Geosains West Science* (Vol. 1, Issue 02).
- Rivanda, A. (2015). *Andrian Rivanda I Pengaruh Paparan Karbon Monoksida Terhadap Daya Konduksi Trakea Majority* (Vol. 4).
- Ranstam, J., Cook, J.A. (2018). LASSO regression. *British Journal of Surgery*. Volume 105, Issue 10, September 2018, Page 1348, <https://doi.org/10.1002/bjs.10895>
- Sihombing, C., Saputra, A. H., Sari, F. P., Mulya, A., Meteorologi, P. S., Indonesia, S., & Artikel, S. (2022). *Prediksi Curah Hujan di Wilayah DKI Jakarta dengan Model NeuralProphet A B S T R A K*. <https://jurnal.stmkg.ac.id/index.php/jam>
- Suryoprayogo, H., Rahmat Iskandar, A., & Adidrana, D. (n.d.). Spatio-Temporal Analysis Polutan Karbon Monoksida (CO) Jakarta Selama Pandemi Menggunakan Sentinel-5P TROPOMI. *DES 2022 Journal of Informatics and Communications Technology*, 41(2), 47–054. <https://doi.org/10.52661>
- SYAIFULLOH, M. M. (2021). PREDIKSI INDEKS STANDAR PENCEMARAN UDARA DI KOTA SURABAYA BERDASARKAN KONSENTRASI GAS KARBON MONOKSIDA. *Jambura Journal of Probability and Statistics*, 2(2), 86–95. <https://doi.org/10.34312/jjps.v2i2.11326>
- Triebe, O., Hewamalage, H., Pilyugina, P., Laptev, N., Bergmeir, C., & Rajagopal, R. (2021). *NeuralProphet: Explainable Forecasting at Scale*
- Uniejewski, B. Weron, R. (2018). *Efficient Forecasting of Electricity Spot Prices with Expert and LASSO Models*. *Energies*. 11(8), 2039. <https://doi.org/10.3390/en11082039>
- Xin, X., Xie, B., & Xiao, Y. (2022). *Robust Fused Lasso Penalized Huber Regression with Nonasymptotic Property and Implementation Studies*. <http://arxiv.org/abs/2209.13281>

- Yermal, L., and Balasubramanian, P., "Application of Auto ARIMA Model for Forecasting Returns on Minute Wise Amalgamated Data in NSE," 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Coimbatore, India, 2017, pp. 1-5, doi: 10.1109/ICCIC.2017.8524232.
- Z. Zhang, "Microsoft Malware Prediction Using LightGBM Model," 2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Xi'an, China, 2022, pp. 41-44, doi: 10.1109/ICBAIE56435.2022.9985850.