

PERBANDINGAN MODEL PREDIKSI DATA MINING DALAM MEMPREDIKSI KONSENTRASI POLUTAN KARBON MONOKSIDA (CO) DI JAKARTA

Rendy Syahril Amanu ^{a,1}, Faiz Ahza Ramadhan ^{b,2,*}, Agung Hari Saputra ^{c,3}

^{a b c} Program Studi Meteorologi, Sekolah Tinggi Meteorologi Klimatologi dan Geofisika, Jalan Perhubungan I No. 5, Kelurahan Pondok Betung, Kecamatan Pondok Aren, 15221, Tangerang Selatan

¹ rendy.syahril17@gmail.com; ² faizahzaramadhan@gmail.com *; ³ agung.hs@stmkg.ac.id

* corresponding author

ARTICLE INFO

Keywords

Pemodelan
Prediksi
Karbon monoksida
Data mining

ABSTRACT

DKI Jakarta, as the capital of Indonesia, faces serious challenges in terms of air quality. Carbon monoxide (CO) is one of the main air pollutants in Jakarta that is harmful to human health and the environment. Data mining is a method that can be used to predict situations based on a model. The study aims to compare data mining models with the best-performing methods to predict carbon monoxide pollutants in Jakarta. The predictive data mining model of the python library is tested and evaluated based on the evaluation metrics of MASE, RMSSE, MAE, RMSE, MAPE and SMAPE values. The model test results showed that K Neighbors with the Conditional Deseasonalize & Detrending model had the best metric evaluation value to predict CO concentration with the value evaluation metrics of MASE 0.2942, RMSSE 0.2483, MAE 2.7362, RMSE 3.3863, MAPE 0.1975 and SMAPE 0.01993. Overall, K Neighbors with the Conditional Deseasonalize & Detrending model shows good performance to predict CO concentrations in Jakarta, but further adjustments are needed to improve accuracy.

1. Pendahuluan

Karbon monoksida (CO) merupakan sebuah gas yang tidak berwarna dan tidak berbau, namun gas karbon monoksida dapat berbahaya jika dihirup dalam jumlah besar [1]. Peningkatan konsentrasi CO di atmosfer berdampak buruk pada kesehatan manusia [2], Selain itu polutan CO juga berkontribusi terhadap terjadinya pemanasan global [3]. Menurut IQAir [4] DKI Jakarta menjadi salah satu dengan kualitas udara tidak sehat. Karbon monoksida menjadi salah satu dari lima parameter dalam Indeks Pencemaran Kualitas Udara [2]. Hal ini menjadikan karbon monoksida sebagai salah satu polutan yang menjadi perhatian karena potensi bahayanya terhadap kesehatan manusia dan lingkungan [5].

Berdasarkan permasalahan kualitas udara Dinas Lingkungan Hidup DKI Jakarta memerlukan suatu pengolahan data yang dapat memberikan informasi lebih banyak [2]. Teknik *data mining* atau biasa lebih dikenal sebagai teknik *machine learning* merupakan sebuah metode yang dapat digunakan untuk memprediksi situasi berdasarkan model prediksi terhadap data yang telah ada [6], *data mining* dapat mengetahui informasi yang lebih rinci berdasarkan data dengan jumlah yang besar[7].

Prediksi *data mining* melibatkan pengolahan data menggunakan model untuk memperoleh pola dari data. [8]. *Data mining* dapat menggali nilai tambah dari suatu data yang tidak dapat diperoleh secara manual [9]. Dengan mengolah data, teknik *data mining* dapat memberikan hasil prediksi dari data tersebut [10].

Dalam beberapa tahun terakhir perkembangan teknologi membuat mengolah data dengan teknik *data mining* memberikan kemudahan pengguna untuk melakukan penelitian[3]. Riset tentang teknik *data mining* semacam ini telah umum dilakukan seperti misalnya riset A. Amalia dkk (2022) [11] yang

melakukan penelitian prediksi kualitas udara menggunakan metode *data mining* atau penelitian A. D. Wiranata dkk (2023) [5] yang melakukan penelitian klasifikasi *data mining* kualitas udara menggunakan model *K-Nearest Neighbors*, serta banyak penelitian lainnya.

Berdasarkan penjelasan diatas, penelitian ini akan membandingkan model *data mining* dengan performa terbaik untuk melakukan prediksi polutan karbon monoksida di Jakarta. Diharapkan hasil penelitian ini dapat menjadi referensi yang berguna bagi pembuat kebijakan dalam melakukan strategi mitigasi polusi udara di Jakarta.

2. Metodologi Penelitian

A. Pengumpulan data

Indeks Standar Pencemaran Udara (ISPU) memberikan gambaran kondisi kualitas udara ambien di suatu lokasi dan waktu tertentu. Pemantauan ISPU berdasarkan tujuh parameter pencemar udara, yaitu Partikulat (PM10 dan PM2.5), Karbon Monoksida (CO), Sulfur Dioksida (SO₂), Nitrogen Dioksida (NO₂), Ozon (O₃), dan Hidrokarbon (HC) [12].

Dalam penelitian ini digunakan data ISPU parameter karbon monoksida (CO) di Jakarta. Di Jakarta, terdapat lima stasiun pemantau kualitas udara (SPKU) yang berlokasi di Bundaran Hotel Indonesia, Jakarta Pusat; Ruang Publik Terpadu Ramah Anak Kelapa Nias III, Jakarta Utara; Area Kebon Bibit Dinas Pertamina Jagakarsa, Jakarta Selatan; Parkir Monumen Lubang Buaya, Jakarta Timur; dan Jalan Jeruk Kuning BI Srengseng, Jakarta Barat [7].

Karbon monoksida adalah salah satu dari tujuh parameter pencemar udara yang dipantau dalam ISPU. Dalam penelitian ini digunakan data ISPU dengan parameter karbon monoksida (CO) dengan rentang waktu dari tahun 2010 hingga 2021. Dengan rentang waktu tersebut, dataset ini memberikan gambaran tentang tren dan pola konsentrasi polutan CO di Jakarta [3].

Dataset Indeks Standar Pencemaran Udara (ISPU) dapat diunduh dari situs web Jakarta Open Data, yang dapat diakses di <https://data.jakarta.go.id/dataset>. Dataset ISPU ini mencakup data pengukuran bulanan tujuh parameter pencemar udara, termasuk karbon monoksida (CO), yang telah dilakukan di lima stasiun pemantau kualitas udara (SPKU) di Jakarta [13]. Dengan contoh data yang telah diunduh.

Tabel 1. Contoh data situs web Jakarta Open Data

tanggal	pm10	so2	co	o3	no2	max	critical	kategori	lokasi_spku
1/1/2010	60	4	73	27	14	73	CO	SEDANG	DKI1
1/2/2010	32	2	16	33	9	33	O3	BAIK	DKI1
1/3/2010	27	2	19	20	9	27	PM10	BAIK	DKI1
...

B. Pra-pemrosesan data

Dataset pada situs web Jakarta Open Data memberikan data bulanan dari tujuh parameter sehingga diperlukan pra-pemrosesan data agar data dapat digunakan dalam melakukan pemodelan. Pra-pemrosesan data untuk dataset diantaranya adalah penyatuan data (*data integration*) dan penanganan nilai yang hilang (*missing values*) [14].

Data integration adalah proses menggabungkan data dari berbagai sumber menjadi satu set data yang konsisten. *Data integration* melibatkan pengumpulan, pemrosesan, dan penyimpanan data dalam satu lokasi [15]. Dari data ISPU yang telah diunduh digunakan nilai pengukuran parameter karbon monoksida (CO) dan tanggal yang disatukan dalam satu kolom sehingga memberikan data deret waktu (*time series*) yang runtut yang tersimpan. Dengan contoh kolom seperti pada Tabel 2.

Tabel 2. Contoh format dataset karbon monoksida

	tanggal	co
1
2

Dalam penelitian ini, metode pra-pemrosesan *data mining* akan digunakan untuk memeriksa dan menangani nilai yang hilang (*missing values*) dalam dataset karbon monoksida (CO). *Missing value* adalah nilai yang hilang dalam dataset yang dapat muncul karena berbagai alasan, seperti kesalahan dalam pengumpulan data atau kegagalan sistem [16].

Untuk menangani *missing value*, Ada beberapa strategi yang dapat digunakan, salah satunya seperti yang digunakan dalam penelitian ini adalah mengisi *missing value* dengan menghitung rata-rata dan kemudian menggantikan *missing value* dengan nilai rata-rata tersebut, yang dikenal sebagai imputasi *mean* [17].

Imputasi *mean* dilakukan dengan menghitung nilai rata-rata (*mean*) untuk variabel *missing values*. Menghitung *mean* hanya dari nilai-nilai yang ada, mengabaikan *missing values*. Selanjutnya data *missing value* diganti dengan nilai *mean* yang telah dihitung. Nilai *mean* memberikan estimasi yang masuk akal untuk *missing value* berdasarkan distribusi data. Dengan *mean* diberikan oleh [18].

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

\bar{x} = nilai rata-rata

x_i = Data ke-i

n = Jumlah data

i = Data ke satu sampai n

Pada tahap pra-pemrosesan data penyatuan data (*data integration*) dan penanganan nilai yang hilang (*missing values*). Penyatuan data dan input nilai *mean* pada nilai yang hilang (*missing values*) pada dataset. Sehingga pada dataset yang telah siap memiliki data dengan jumlah 4383 baris konsentrasi polutan karbon monoksida (CO). Tabel 3 adalah contoh untuk dataset yang telah siap.

Tabel 3. Dataset yang siap diolah

	tanggal	co
1	01/01/2010	73
2	02/01/2010	16
3	03/01/2010	19
...
4381	29/12/2021	15
4382	30/12/2021	17
4383	31/12/2021	44

C. Konfigurasi Pengujian Model

Untuk melakukan percobaan prediksi menggunakan *library python data mining*, digunakan sebuah *library* PyCaret. *Library* PyCaret adalah sebuah *library low code* untuk sebuah *machine learning*. *Low code* berarti hanya membutuhkan sedikit kode dalam penggunaannya [16]. Selain itu *library* PyCaret dapat melakukan komparasi beberapa model untuk mendapatkan performa model terbaik [19].

Sebelum melakukan pengujian model, dilakukan konfigurasi untuk mengatur lingkungan percobaan dalam *library* PyCaret [6]. Dalam penelitian ini digunakan konfigurasi *library* PyCaret seperti yang dijelaskan pada Tabel 4.

Tabel 4. Konfigurasi *setup data mining*

Description	Setup parameter	value
Target	target	co
forecast horizon	fh	31
session_id	session_id	123
Fold Number	fold	5
level of output	verbose	TRUE
Remove Harmonics	remove_harmonics	TRUE
Seasonal Period(s) Tested	seasonal_period	7
Fold Generator	fold_strategy	sliding
Max Period to Consider	max_sp_to_consider	48
Harmonics Order Method	harmonic_order_method	harmonic_strength
Num Seasonalities to Use	num_sps_to_use	1
confidence level	coverage	0,7

D. Pengujian Model Prediksi

Untuk mendapatkan model *data mining* terbaik dari *library* python, dilakukan pengujian menggunakan *library* PyCaret [6]. Dataset karbon monoksida (CO) yang telah siap diproses menggunakan model yang berada dalam *library* PyCaret, Seperti yang dijelaskan pada Tabel 5 [20].

Tabel 5. Model prediksi

Index	Model	Keunggulan	Kekurangan	Refrensi
ada_cds_dt	AdaBoost	Fleksibel dan optimalisasi model	Sensitif pada fitur yang banyak	[21][22]
arima	ARIMA	Fleksibilitas	Rumit dan waktu proses yang lama	[23][24]
auto_arima	Auto ARIMA	Fleksibilitas	Waktu proses yang lama	[19][24]
br_cds_dt	Bayesian Ridge	Adaptasi model dan pelatihan yang cepat	Kompleksitas komputasi	[25]
croston	Croston	Sederhana dan Efisien	Bias dalam Peramalan	
dt_cds_dt	Decision Tree	Mengolah data dalam berbagai format	Sensitif terhadap noise.	[9][26]
en_cds_dt	Elastic Net	Menangani kekurangan model Lasso	Kompleksitas model	[27]
et_cds_dt	Extra Trees	Efisien	Sensitivitas terhadap Noise	[28]
xgboost_cds_dt	Extreme Gradient Boosting	Penanganan missing value	Membutuhkan waktu untuk pelatihan	[29]
gbr_cds_dt	Gradient Boosting	Penanganan missing value	Membutuhkan waktu untuk pelatihan	[29]
grand_means	Grand Means Forecaster	Sederhana untuk dipahami dan diimplementasikan	Tidak menangkap pola kompleks dalam data	[30]
huber_cds_dt	Huber	Menangani data outlier lebih baik	Parameter yang tepat untuk hasil optimal	[31]
knn_cds_dt	K Neighbors	Menangani data berskala besar	Sensitif terhadap data noise dan outlier	[18]



llar_cds_dt	<i>Lasso Least Angular Regressor</i>	Efisien dalam mengatasi multikolinearitas	Tidak bekerja dengan baik jika ada noise atau outlier dalam data	[32]
lasso_cds_dt	<i>Lasso</i>	Efisien dalam mengatasi multikolinearitas	Tidak bekerja dengan baik jika ada noise atau outlier dalam data	[32]
lightgbm_cds_dt	<i>Light Gradient Boosting</i>	Efisien dalam penggunaan memori dan waktu	Hasil bias jika nilai precision dan recall berbeda jauh dari nilai akurasi	[33]
Index	Model	Keunggulan	Kekurangan	Referensi
lr_cds_dt	<i>Linear</i>	Mudah dipahami dan efisien	Membuat asumsi bahwa hubungan antara variabel bebas	[34]
naive	<i>Naive Forecaster</i>	Mudah dipahami dan efisien	Tidak dapat menangkap pola yang kompleks	[35]
omp_cds_dt	<i>Orthogonal Matching Pursuit</i>	Lebih cepat dan efisien	Tidak bekerja dengan baik jika ada noise atau outlier dalam data	[36][37]
rf_cds_dt	<i>Random Forest</i>	Lebih cepat dan efisien	Akurasi validasinya mungkin lebih rendah	[7][38]
ridge_cds_dt	<i>Ridge</i>	Menangani multicollinearity dan mencegah overfitting	Tidak bekerja dengan baik jika ada outlier dalam data	[39]
snaive	<i>Seasonal Naive Forecaster</i>	Hasil yang cukup akurat jika data memiliki pola musiman yang stabil	Tidak bekerja dengan baik jika ada fluktuasi non-musiman dalam data	[40]
stlf	<i>STLF</i>	Fleksibilitas yang tinggi, lebih modern dibandingkan dengan ARIMA	Dipengaruhi banyaknya data hilang, hanya sesuai untuk beberapa jenis data	[41]

E. Dekomposisi Data Untuk Prediksi

Untuk mendapatkan analisis dataset karbon monoksida digunakan metode dekomposisi *addictive*. Metode tersebut digunakan untuk pemecahan data aktual guna mendapatkan analisis data konsentrasi CO di Jakarta [42]. Metode dekomposisi *addictive* digunakan dalam analisis data *time series* yang memecah data menjadi beberapa pola dan mengidentifikasi masing-masing komponen dari data tersebut secara terpisah untuk melakukan mendukung model prediksi [43]. Dalam penelitian ini digunakan metode dekomposisi untuk pemisahan data menjadi komponen tren, musiman, dan residu yang kemudian dianalisis untuk mendukung model prediksi [44].

F. Evaluasi Prediksi Model Data Mining

Untuk mendapatkan model prediksi konsentrasi karbon monoksida terbaik dari *library data mining*[6]. Model dalam *library* python diuji untuk melihat keakuratan model memprediksi konsentrasi karbon monoksida berdasarkan data historis [45]. Menggunakan subset pengujian, evaluasi model data mining dengan metrik yang sesuai [46].

Setiap model diuji menggunakan sejumlah metrik evaluasi, termasuk MASE (*Mean Absolute Scaled Error*), RMSSE (*Root Mean Squared Scaled Error*), MAE (*Mean Absolute Error*), RMSE (*Root Mean Squared Error*), MAPE (*Mean Absolute Percentage Error*) dan SMAPE (*Symmetric Mean Absolute Percentage Error*) [47]. Metrik evaluasi tersebut memberikan ukuran tentang seberapa baik model dapat memprediksi konsentrasi karbon monoksida, dengan nilai yang lebih rendah menunjukkan kinerja yang lebih baik [48].



Model dengan nilai evaluasi terbaik, yang menunjukkan tingkat kesalahan prediksi terendah dan kemampuan penjelasan variabilitas data yang baik, dipilih sebagai model terbaik untuk digunakan dalam prediksi konsentrasi CO di Jakarta [11].

Model prediksi terbaik kemudian di analisis berdasarkan nilai metrik evaluasi MASE (*Mean Absolute Scaled Error*), RMSSE (*Root Mean Squared Scaled Error*), MAE (*Mean Absolute Error*), RMSE (*Root Mean Squared Error*), MAPE (*Mean Absolute Percentage Error*) dan SMAPE (*Symmetric Mean Absolute Percentage Error*) [47]. Nilai model terbaik di evaluasi pada tiap nilai metrik untuk melihat performa model [49]. Nilai metrik evaluasi model diperoleh berdasarkan perhitungan berikut.

MASE (*Mean Absolute Scaled Error*) diberikan oleh [50]:

$$MASE = \frac{1}{h} \sum_{j=1}^h \frac{|y_{t+j} - \hat{y}_{t+j}|}{\Delta_y} \quad (2)$$

RMSSE (*Root Mean Squared Scaled Error*) diberikan oleh [51]:

$$RRMSE = \sqrt{\frac{1}{h} \sum_{j=1}^h \frac{[(y_{t+j}) - (\hat{y}_{t+j})]^2}{\Delta_y^2}} \quad (3)$$

MAE (*Mean absolute error*) diberikan oleh [52]:

$$MAE = \frac{1}{h} \sum_{j=1}^h (y_{t+j} - \hat{y}_{t+j}) \quad (4)$$

RMSE (*Root Mean Squared Error*) diberikan oleh [52]:

$$RMSE = \sqrt{\frac{1}{h} \sum_{j=1}^h (y_{t+j} - \hat{y}_{t+j})^2} \quad (5)$$

MAPE (*Mean Absolute Percentage Error*) diberikan oleh [53]

$$MAPE = \frac{1}{h} \sum_{j=1}^h \frac{|y_{t+j} - \hat{y}_{t+j}|}{y_{t+j}} \quad (6)$$

SMAPE (*Symmetric Mean Absolute Percentage Error*) diberikan oleh [53]:

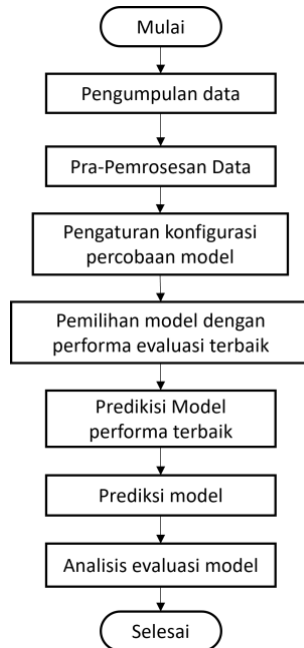
$$SMAPE = \frac{1}{h} \sum_{j=1}^h \frac{[(y_{t+j}) - (\hat{y}_{t+j})]}{\frac{[(\hat{y}_{t+j}) + (y_{t+j})]}{2}} \quad (7)$$

Keterangan : y_{t+j} = data aktual ke-i,
 \hat{y}_{t+j} = nilai prediksi ke-i, dan
 h = jumlah data.

G. Implementasi Model Prediksi

Model terbaik dalam melakukan pengujian pemodelan *forecast* polutan CO kemudian di implementasikan dalam melakukan *forecast* atau prediksi polutan CO di Jakarta. Dengan melakukan *plotting* hasil prediksi menggunakan grafik untuk selanjutnya dilakukan analisis [54].

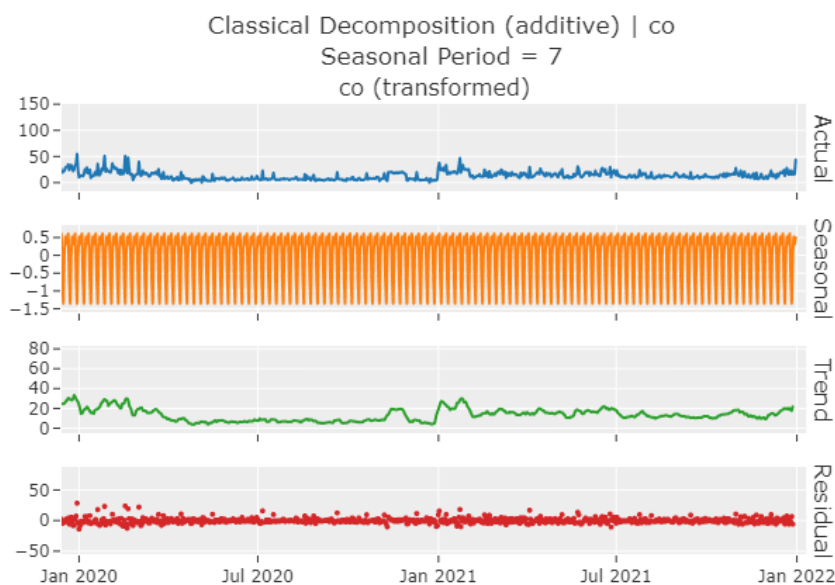
Dalam penelitian ini menggunakan tahapan penelitian seperti dijelaskan pada gambar 1.



Gambar 1. Tahap pengolahan data

3. Hasil dan Pembahasan

A. Dekomposisi Data Aktual



Gambar 2. Dekomposisi dataset karbon monoksida

Seperti yang ada pada gambar 2, dalam analisis *time series* model prediksi memungkinkan untuk memecah data *time series* menjadi beberapa komponen. Seperti pada gambar 2 merupakan hasil dekomposisi dataset CO, dataset CO dipecah menggunakan dekomposisi klasik dengan metode aditif pada data *time series* untuk konsentrasi karbon monoksida (CO) [55]. Pada hasil dekomposisi data aktual menampilkan grafik *Seasonal* (musiman), *Trend* (tren), *Residual* (residu). Secara umum grafik tersebut berguna untuk memberikan pemahaman dari faktor serta gangguan yang kemungkinan berpengaruh pada proses prediksi konsentrasi polutan CO.

Dari grafik trend, menggambarkan bahwa tren tersebut relatif stabil dengan beberapa grafik fluktuatif. Pada Januari 2020 hingga Maret 2020 terlihat tren konsentrasi polutan karbon monoksida mengalami tren yang menurun. Setelah itu, tren menunjukkan stabilisasi dengan fluktuasi kecil hingga akhir tahun 2020. Pada rentang waktu Januari 2021 hingga Desember 2021 terlihat grafik tren sedikit terjadi peningkatan dengan beberapa fluktuasi konsentrasi CO. Secara umum grafik tren menunjukkan bahwa konsentrasi CO di Jakarta memiliki kecenderungan yang stabil, dengan beberapa fluktuasi yang terjadi seperti pada sekitar Januari 2020 dan sekitar Januari 2021.

B. Hasil Pengujian Model Prediksi

Nilai MAE dan RMSE yang lebih rendah menunjukkan bahwa model ini memiliki kesalahan absolut dan kuadrat rata-rata yang lebih rendah, yang berarti model ini lebih akurat dalam memprediksi data. Berdasarkan Tabel 6 hasil pengujian beberapa model *data mining*, menunjukkan bahwa model *K Neighbors with Conditional Deseasonalize & Detrending* memiliki nilai MAE dan RMSSE terendah dengan nilai MAE sebesar 2.739 dan nilai RMSE sebesar 3.3863, dimana nilai evaluasi model model *K Neighbors with Conditional Deseasonalize & Detrending* tersebut lebih rendah dibandingkan model *data mining* lainnya. Model *K Neighbors with Conditional Deseasonalize & Detrending* juga memiliki nilai terendah pada metrik evaluasi MASE dan RMSSE. Model tersebut memiliki nilai MASE sebesar 0.2942 dan nilai RMSSE sebesar 0.2483

Selain itu, nilai MAPE dan SMAPE yang lebih rendah menunjukkan bahwa model ini memiliki kesalahan persentase yang lebih rendah. Pada hasil pengujian beberapa model prediksi model *K Neighbors with Conditional Deseasonalize & Detrending* memiliki nilai MAPE dan SMAPE yang lebih rendah daripada model lainnya. Nilai MAPE model *K Neighbors with Conditional Deseasonalize & Detrending* sebesar 0.1975 dan nilai SMAPE sebesar 0.1993, dimana nilai evaluasi model *K Neighbors with Conditional Deseasonalize & Detrending* lebih kecil daripada model lainnya. Sehingga model *K Neighbors with Conditional Deseasonalize & Detrending* memiliki kesalahan presentase yang lebih rendah daripada model lainnya.

Berdasarkan penjelasan tersebut serta hasil pengujian model prediksi yang disajikan pada Tabel 4, model *K Neighbors with Conditional Deseasonalize & Detrending* sebagai model peramalan dengan performa terbaik karena memiliki nilai akurasi terbaik untuk metrik evaluasi MASE, RMSSE, MAE, RMSE, MAPE dan SMAPE dengan waktu proses model *K Neighbors with Conditional Deseasonalize & Detrending* selama 0.3720 detik. Sebab nilai evaluasi menunjukkan bahwa model *K Neighbors* memiliki tingkat kesalahan yang lebih rendah dibandingkan model lainnya dalam tabel.

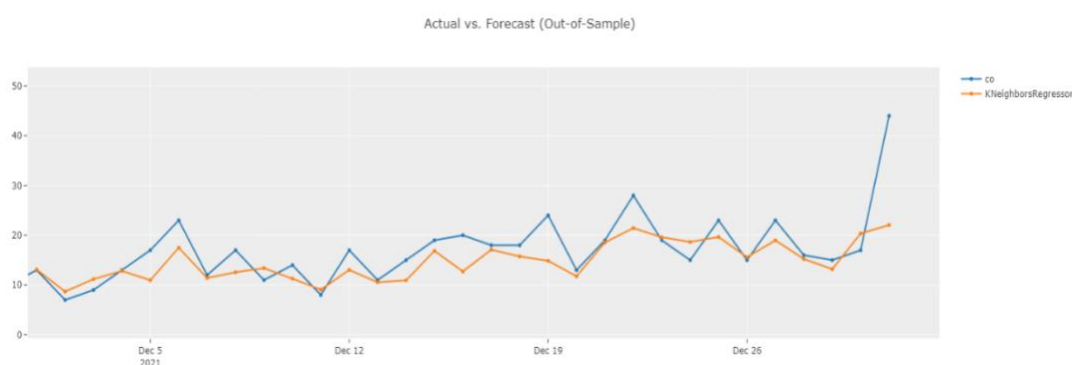
Tabel 6. Model dengan performa prediksi terbaik

Model	MASE	RMSSE	MAE	RMSE	MAPE	SMAPE	TT (Sec)
<i>K Neighbors with Cond. Deseasonalize & Detrending</i>	0.2942	0.2483	2.7362	3.3863	0.1975	0.1993	0.3720
<i>Linear with Cond. Deseasonalize & Detrending</i>	0.3400	0.2954	3.1630	4.0291	0.2360	0.2215	0.4340



<i>Ridge with Cond. Deseasonalize & Detrending</i>	0.3400	0.2954	3.1630	4.0291	0.2360	0.2215	0.4180
<i>Bayesian Ridge with Cond. Deseasonalize & Detrending</i>	0.3404	0.2953	3.1665	4.0275	0.2368	0.2220	0.3080
<i>Elastic Net with Cond. Deseasonalize & Detrending</i>	0.3414	0.2958	3.1761	4.0349	0.2379	0.2229	0.3020
<i>Light Gradient Boosting with Cond. Deseasonalize & Detrending</i>	0.3421	0.2812	3.1816	3.8348	0.2654	0.2344	15.360
Model	MASE	RMSSE	MAE	RMSE	MAPE	SMAPE	TT (Sec)
<i>Lasso Least Angular Regressor with Cond. Deseasonalize & Detrending</i>	0.3427	0.2964	3.1882	4.0424	0.2395	0.2242	0.4640
<i>Lasso with Cond. Deseasonalize & Detrending</i>	0.3427	0.2964	3.1881	4.0424	0.2395	0.2241	0.4780
<i>Huber with Cond. Deseasonalize & Detrending</i>	0.3542	0.2996	3.2943	4.0859	0.2342	0.2350	0.3780
<i>Auto ARIMA</i>	0.3738	0.3761	3.4780	5.1303	0.2757	0.2626	2.790.880
<i>Random Forest with Cond. Deseasonalize & Detrending</i>	0.3892	0.3200	3.6208	4.3648	0.3198	0.2719	45.980
<i>Extreme Gradient Boosting with Cond. Deseasonalize & Detrending</i>	0.4073	0.3297	3.7885	4.4965	0.3040	0.2611	0.6220
<i>Gradient Boosting with Cond. Deseasonalize & Detrending</i>	0.4175	0.3399	3.8846	4.6366	0.3538	0.2916	20.040
<i>Extra Trees with Cond. Deseasonalize & Detrending</i>	0.4349	0.3382	4.0460	4.6131	0.3437	0.2975	15.000
<i>Croston</i>	0.4882	0.3897	4.5419	5.3150	0.3579	0.3356	0.0420
<i>ARIMA</i>	0.4896	0.5529	4.5543	7.5401	0.3689	0.3319	13.360
<i>STLF</i>	0.5346	0.4287	4.9732	5.8467	0.3870	0.3741	0.7480
<i>Decision Tree with Cond. Deseasonalize & Detrending</i>	0.5400	0.4311	5.0231	5.8796	0.3743	0.3766	0.3840
<i>Orthogonal Matching Pursuit with Cond. Deseasonalize & Detrending</i>	0.5520	0.4388	5.1380	5.9871	0.5107	0.3629	0.3000
<i>Naive Forecaster</i>	0.5588	0.4540	5.2000	6.1935	0.4131	0.3866	14.740
<i>Seasonal Naive Forecaster</i>	0.5865	0.4764	5.4571	6.4978	0.4367	0.4160	0.1060
<i>AdaBoost with Cond. Deseasonalize & Detrending</i>	13.436	10.234	12.4980	13.9571	10.679	0.6332	0.6880
<i>Grand Means Forecaster</i>	17.171	12.111	15.9806	16.5236	14.874	0.7630	0.0540

C. Hasil Prediksi Model Terhadap Konsentrasi CO



Gambar 3. Komparasi prediksi model *K Neighbors with Conditional Deseasonalize & Detrending* dengan data aktual

Pada gambar 3, menampilkan dua garis, yaitu *co* dengan warna biru mewakili data aktual, dan garis *K NeighborsRegressor* berwarna oranye mewakili data hasil prediksi dari model *K Neighbors with Conditional Deseasonalize & Detrending*.

Kedua garis menunjukkan tren yang serupa dengan fluktuasi nilai dari waktu ke waktu. Tampak bahwa model *K Neighbors with Conditional Deseasonalize & Detrending* memiliki tingkat akurasi yang cukup baik dalam memprediksi konsentrasi karbon monoksida (CO) untuk sebagian besar periode yang ditampilkan. Garis biru yang mewakili data aktual dan garis oranye yang mewakili prediksi model cenderung mengikuti pola yang sama, naik dan turun secara bersamaan, yang menunjukkan bahwa model tersebut mampu menangkap tren umum dalam data.

Namun, terdapat beberapa perbedaan antara nilai aktual dan prediksi, yang merupakan hal yang umum dalam pemodelan prediktif. Perbedaan ini bisa disebabkan oleh berbagai faktor, seperti perubahan mendadak dalam data yang tidak dapat diprediksi oleh model atau keterbatasan model dalam menangkap semua dinamika yang ada dalam data.

D. Evaluasi Model Forecast Konsentrasi CO

Tabel 7. metrik nilai evaluasi model *K Neighbors with Conditional Deseasonalize & Detrending*

	cutoff	MASE	RMSSE	MAE	RMSE	MAPE	SMAPE
0	6/28/2021	0.3703	0.2842	3.5246	3.9414	0.2898	0.2458
1	7/29/2021	0.3391	0.2655	3.2057	3.6653	0.2932	0.2454
2	8/29/2021	0.2258	0.1944	2.1220	2.6740	0.2068	0.1775
3	9/29/2021	0.4616	0.3882	4.3092	5.3242	0.3560	0.2840
4	10/30/2021	0.3272	0.3099	3.0412	4.2425	0.2470	0.2295
Mean		0.3448	0.2884	3.2405	3.9695	0.2786	0.2364
SD		0.0759	0.0629	0.7092	0.8581	0.0500	0.0345

Nilai MASE yang berfluktuasi menunjukkan adanya variasi dalam kinerja model relatif terhadap model naif. Nilai terendah pada 8/29/2021 sebesar 0.1944 menandakan kinerja terbaik model pada tanggal tersebut. Rata-rata MASE adalah 0.2258, dengan standar deviasi (SD) 0.0759.

Nilai RMSSE yang semuanya di bawah satu menunjukkan bahwa model secara konsisten memiliki kinerja yang baik. Nilai terendah sebesar 0.1944 pada 8/29/2021 menunjukkan kinerja terbaik model. Dengan rata-rata nilai RMSSE adalah 0.2884, dengan SD 0.0629.

Nilai MAE yang bervariasi menunjukkan besarnya rata-rata kesalahan prediksi model. Nilai terendah pada tanggal 8/29/2021 sebesar 2.1220 menandakan kesalahan rata-rata terkecil pada tanggal tersebut. Dengan rata-rata nilai MAE adalah 3.2405, dengan SD 0.7092.

Nilai RMSE berfungsi untuk mengukur tingkat kesalahan hasil prediksi dari suatu model. Dimana semakin kecil nilai RMSE, semakin baik performa model. Nilai terendah pada tanggal 8/29/2021 sebesar 2.6740 menunjukkan kinerja terbaik model pada tanggal tersebut. Dengan rata-rata nilai MAE adalah 3.9695, dengan SD 0.8581.

Nilai MAPE yang lebih rendah menunjukkan akurasi prediksi yang lebih tinggi sebagai persentase. Nilai terendah sebesar 0.2068 pada tanggal 8/29/2021 menunjukkan kinerja terbaik model pada tanggal tersebut. Rata-rata MAPE adalah 0.2786, dengan SD 0.0500.

Nilai SMAPE yang lebih rendah menunjukkan akurasi prediksi yang lebih tinggi dengan memperlakukan *overestimate* dan *underestimate* secara setara. Nilai terendah dengan nilai 0.1775 pada 8/29/2021 menunjukkan kinerja terbaik model pada tanggal tersebut. Rata-rata SMAPE adalah 0.2364, dengan SD 0.0345.

Berdasarkan nilai metrik MASE, RMSSE, MAE, RMSE, MAPE dan SMAPE model *K Neighbors with Conditional Deseasonalize & Detrending* memiliki nilai evaluasi metrik dengan tingkat kesalahan yang rendah. Hal ini menunjukkan model *K Neighbors with Conditional Deseasonalize & Detrending* memiliki kemampuan yang baik dalam menjelaskan variabilitas data dan melakukan prediksi dengan tingkat kesalahan yang rendah. Dengan demikian dapat disimpulkan bahwa model tersebut memiliki performa yang baik dalam memprediksi polutan karbon monoksida di Jakarta.

4. Kesimpulan

Berdasarkan analisis data dan evaluasi model, dapat disimpulkan bahwa model *K Neighbors with Conditional Deseasonalize & Detrending* adalah model dengan performa terbaik dalam memprediksi konsentrasi karbon monoksida (CO) berdasarkan dataset yang diberikan. Model ini memproses data selama 0.3720 detik dan menunjukkan tingkat akurasi yang lebih tinggi dibandingkan model lainnya dengan nilai MASE sebesar 0.2942, RMSSE sebesar 0.2483, MAE sebesar 2.7362, RMSE sebesar 3.3863, MAPE sebesar 0.1975 dan SMAPE sebesar 0.1993.

Hasil prediksi model *K Neighbors with Conditional Deseasonalize & Detrending* menunjukkan bahwa model ini mampu menangkap tren umum dalam data dan memberikan prediksi yang cukup akurat untuk sebagian besar periode yang ditampilkan. Namun, ada beberapa perbedaan antara nilai aktual dan prediksi, yang mungkin disebabkan oleh berbagai faktor. Secara keseluruhan, model *K Neighbors with Conditional Deseasonalize & Detrending* ini dapat digunakan untuk memprediksi konsentrasi CO, tetapi perlu dilakukan penyesuaian dan peningkatan untuk meningkatkan akurasi

Daftar Pustaka

- [1] D. Y. Damara, I. W. Wardhana, dan E. Sutrisno, "ANALISIS DAMPAK KUALITAS UDARA KARBON MONOKSIDA (CO) DI SEKITAR JL. PEMUDA AKIBAT KEGIATAN CAR FREE DAY MENGGUNAKAN PROGRAM CALINE4 DAN SURFER (STUDI KASUS: KOTA SEMARANG)," *Jurnal Teknik Lingkungan*, vol. 6, no. 1, 2017.
- [2] A. I. Sang, E. Sutoyo, dan I. Darmawan, "ANALISIS DATA MINING UNTUK KLASIFIKASI DATA KUALITAS UDARA DKI JAKARTA MENGGUNAKAN ALGORITMA DECISION TREE DAN SUPPORT VECTOR MACHINE DATA MINNING ANALYSIS FOR CLASSIFICATION OF AIR QUALITY DATA DKI JAKARTA USING DECISION TREE ALGORITHM AND SUPPORT VECTOR MACHINER ALGORITHM," *e-Proceeding of Engineering*, vol. 8, no. 5, hlm. 8954–8963, 2021.
- [3] M. Astriyani, I. N. Laela, D. P. Lestari, L. Anggraeni, dan T. Astuti, "ANALISIS KLASIFIKASI DATA KUALITAS UDARA DKI JAKARTA MENGGUNAKAN ALGORITMA C.45," *JURNAL SISTEM & TEKNOLOGI INFORMASI KOMUNIKASI*, vol. 6, no. 1, hlm. 36–41, 2022.
- [4] IQAir, "2019 WORLD AIR QUALITY REPORT Region & City PM2.5 Ranking," 2019.
- [5] A. D. Wiranata, S. Soleman, I. Irwansyah, I. K. Sudaryana, dan R. Rizal, "KLASIFIKASI DATA MINING UNTUK MENENTUKAN KUALITAS UDARA DI PROVINSI DKI JAKARTA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS (K-NN)," *Infotech: Journal of Technology Information*, vol. 9, no. 1, hlm. 95–100, Jun 2023, doi: 10.37365/jti.v9i1.164.
- [6] A. R. Abidin, I. Kadek, dan D. Nuryana, "Perbandingan Metode Klasifikasi Data Mining Untuk Mengukur Tingkat Kepuasan Mahasiswa Terhadap Sistem Informasi Penilaian Nonakademik UNESA (SIPENA)," *Journal of Emerging Information Systems and Business Intelligence*, vol. 04, no. 04, hlm. 129–138, 2023.



- [7] A. Nugroho, I. Asror, dan Y. F. A. Wibowo, "Klasifikasi Tingkat Kualitas Udara DKI Jakarta Berdasarkan Open Government Data Menggunakan Algoritma Random Forest," *e-Proceeding of Engineering*, vol. 10, No.2, hlm. 1824, Apr 2023.
- [8] D. H. Kamagi dan S. Hansun, "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," *Jurnal ULTIMATICS*, vol. 6, no. 1, hlm. 15–20, Jun 2014, doi: 10.31937/ti.v6i1.327.
- [9] A. Muzakir dan R. A. Wulandari, "Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree," *Scientific Journal of Informatics*, vol. 3, no. 1, 2016, [Daring]. Tersedia pada: <http://journal.unnes.ac.id/nju/index.php/sji>
- [10] S. Qomariah, H. Ekawati, dan S. Belareq, "IMPLEMENTASI METODE DATA MINING APRIORI PADA APLIKASI PENJUALAN PT. TIGA RAKSA SATRIA," *Komputasi: Jurnal Ilmiah Ilmu Komputer dan Matematika*, vol. 17, no. 1, hlm. 329–338, Jan 2020, doi: 10.33751/komputasi.v17i1.1747.
- [11] A. Amalia, A. Zaidiah, dan I. N. Isnainiyah, "PREDIKSI KUALITAS UDARA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR," *Jurnal Ilmiah Penelitian dan Pembelajaran Informatika*, vol. 07, no. 02, hlm. 465–507, Jun 2022, [Daring]. Tersedia pada: <https://data.jakarta.go.id/>.
- [12] A. Nugroho, I. Asror, dan Y. F. A. Wibowo, "Klasifikasi Tingkat Kualitas Udara DKI Jakarta Berdasarkan Open Government Data Menggunakan Algoritma Random Forest," *e-Proceeding of Engineering*, vol. 10, no. 2, hlm. 1824–1834, 2023.
- [13] Portal Data Terpadu Pemprov DKI Jakarta, "Open Data Jakarta." Diakses: 29 Desember 2023. [Daring]. Tersedia pada: <https://data.jakarta.go.id/>
- [14] R. F. Putra *dkk.*, *DATA MINING: Algoritma dan Penerapannya*. PT. Sonpedia Publishing Indonesia, 2023.
- [15] G. K. Geofanny, "Sistem Manajemen Risiko Berbasis ISO 31000: 2018 Di PT. Bawen Mediatama," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 4, hlm. 2870–2878, 2022.
- [16] M. T. Anwar dan D. R. A. Permana, "Perbandingan Performa Model Data Mining untuk Prediksi Dropout Mahasiswa," *Jurnal Teknologi dan Manajemen*, vol. 19, no. 2, hlm. 33–40, Agu 2021, doi: 10.52330/jtm.v19i2.34.
- [17] M. D. Purbolaksono, M. Irvan Tantowi, A. Imam Hidayat, dan A. Adiwijaya, "Perbandingan Support Vector Machine dan Modified Balanced Random Forest dalam Deteksi Pasien Penyakit Diabetes," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 2, hlm. 393–399, Apr 2021, doi: 10.29207/resti.v5i2.3008.
- [18] R. M. Syauqi, P. N. Sabrina, dan I. Santikarama, "K-Means Clustering with KNN and Mean Imputation on CPU Benchmark Compilation Data," *Journal of Applied Informatics and Computing*, vol. 7, no. 2, hlm. 231–239, Des 2023, doi: 10.30871/jaic.v7i2.6491.
- [19] M. T. Anwar dan D. R. A. Permana, "Perbandingan Performa Model Data Mining untuk Prediksi Dropout Mahasiswa," *Jurnal Teknologi dan Manajemen*, vol. 19, no. 2, hlm. 33–40, Agu 2021, doi: 10.52330/jtm.v19i2.34.
- [20] Moez Ali, "PyCaret: An open source, low-code machine learning library in Python," PyCaret version 1.0. Diakses: 3 Januari 2024. [Daring]. Tersedia pada: <https://www.pycaret.org>
- [21] A. Byna dan M. Basit, "PENERAPAN METODE ADABOOST UNTUK MENGOPTIMASI PREDIKSI PENYAKIT STROKE DENGAN ALGORITMA NAÏVE BAYES," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 9, no. 3, hlm. 407–411, Nov 2020, doi: 10.32736/sisfokom.v9i3.1023.
- [22] L. D. Utami, S. Tinggi, M. Informatika, D. Komputer, N. Mandiri, dan R. S. Wahono, "Integrasi Metode Information Gain Untuk Seleksi Fitur dan Adaboost Untuk Mengurangi Bias Pada Analisis Sentimen Review Restoran Menggunakan Algoritma Naïve Bayes," *Journal of Intelligent Systems*, vol. 1, no. 2, 2015, [Daring]. Tersedia pada: <http://journal.ilmukomputer.org>



- [23] A. H. Hutasuht, W. Anggraeni, dan R. Tyasnurita, "Pembuatan Aplikasi Pendukung Keputusan Untuk Peramalan Persediaan Bahan Baku Produksi Plastik Blowing dan Inject Menggunakan Metode ARIMA (Autoregressive Integrated Moving Average) Di CV. Asia," *JURNAL TEKNIK POMITS*, vol. 3, no. 2, hlm. 169–174, 2014.
- [24] S. R. Ningtias, "Aplikasi metode double exponential smoothing holt dan arima untuk meramalkan voluntary counseling and testing (Vct) odha di Provinsi Jawa Timur," *The Indonesian Journal of Public Health*, vol. 13, no. 2, 2018.
- [25] W. A. Septiana dan S. Soehardjoepri, "Penerapan Metode Bayes dalam Menentukan Model Estimasi Reliabilitas Pompa Submersible pada Rumah Pompa Wendit I PDAM Kota Malang," *Jurnal Sains dan Seni ITS*, vol. 6, no. 2, hlm. A18–A23, 2017.
- [26] S. Supangat, A. R. Amna, dan T. Rahmawati, "Implementasi Decision Tree C4.5 Untuk Menentukan Status Berat Badan dan Kebutuhan Energi Pada Anak Usia 7-12 Tahun," *Teknika*, vol. 7, no. 2, hlm. 73–78, Nov 2018, doi: 10.34148/teknika.v7i2.90.
- [27] H. E. Sinaga dan D. R. S. Saputro, "Performa Metode Elastic-Net dalam Kasus Multikolinearitas pada Analisis Linear Berganda," *Prosiding Seminar Pendidikan Matematika dan Matematika*, vol. 3, Mei 2021, doi: 10.21831/pspmm.v3i0.123.
- [28] A. Dharma, P. Manalu, G. S. Sinaga, R. Siringoringo, I. S. Palangai, dan K. Setiawan, "Deteksi Pola Pasien Kanker Serviks dengan Algoritma Extra Trees dan K-Nearest Neighbor," *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, vol. 3, no. 1, hlm. 32–36, 2020.
- [29] N. B. A. B. Arifin dan A. Asmianto, "SISTEM PREDIKSI PENJUALAN MENGGUNAKAN KOMBINASI METODE MONTE CARLO DAN DECISION TREE BERBASIS WEBSITE," *MATHunesa: Jurnal Ilmiah Matematika*, vol. 11, no. 2, hlm. 274–286, Agu 2023, doi: 10.26740/mathunesa.v11n2.p274-286.
- [30] J. C. Avendano, L. D. Otero, dan C. Otero, "Application of Statistical Machine Learning Algorithms for Classification of Bridge Deformation Data Sets," dalam *2021 IEEE International Systems Conference (SysCon)*, IEEE, Apr 2021, hlm. 1–7. doi: 10.1109/SysCon48628.2021.9447056.
- [31] C. M. Liapis dan S. Kotsiantis, "Energy Balance Forecasting: An Extensive Multivariate Regression Models Comparison," dalam *Proceedings of the 12th Hellenic Conference on Artificial Intelligence*, New York, NY, USA: ACM, Sep 2022, hlm. 1–7. doi: 10.1145/3549737.3549782.
- [32] A. P. Andana, D. Safitri, dan A. Rusgionyono, "MODEL REGRESI MENGGUNAKAN LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR (LASSO) PADA DATA BANYAKNYA GIZI BURUK KABUPATEN/KOTA DI JAWA TENGAH," 2017. [Daring]. Tersedia pada: <https://api.semanticscholar.org/CorpusID:125666234>
- [33] F. I. Kurniadi dan P. D. Larasati, "Light Gradient Boosting Machine untuk Deteksi Penyakit Stroke," *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan)*, vol. 6, no. 1, hlm. 67–72, Okt 2022, doi: 10.47970/siskom-kb.v6i1.328.
- [34] F. H. Hamdanah dan D. Fitrihanah, "Analisis Performansi Algoritma Linear Regression dengan Generalized Linear Model untuk Prediksi Penjualan pada Usaha Mikra, Kecil, dan Menengah," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 10, no. 1, hlm. 23, Apr 2021, doi: 10.23887/janapati.v10i1.31035.
- [35] S. Afralia Hevit, T. Fahrudin, dan Rochmawati, "PERBANDINGAN MODEL ESTIMASI INFLASI PANGAN MENGGUNAKAN ALGORITMA NAIVE FORECASTER DAN EXPONENTIAL SMOOTHING," *Jurnal Manajemen dan Akuntansi*, vol. 18, no. 2, hlm. 566–580, 2023.
- [36] A. Fatahillah dan I. S. Winardhi, "Perbandingan Inversi Reflektifitas Menggunakan Basis Pursuit Inversion dan Orthogonal Matching Pursuit pada Lapisan Tipis," *Jurnal Geofisika*, vol. 15, no. 2, hlm. 1, Nov 2019, doi: 10.36435/jgf.v15i2.401.



- [37] S. Maria, "Extension de Matching Pursuit au modèle des moindres carrés totaux," 2005. [Daring]. Tersedia pada: <https://api.semanticscholar.org/CorpusID:169460737>
- [38] I. Wardhana, M. Ariawijaya, V. A. Isnaini, R. P. Wirman, dan S. T. S. Jambi, "Media Elektronik : 2580-0760 Gradient Boosting Machine , Random Forest dan Light," 2022. [Daring]. Tersedia pada: <https://api.semanticscholar.org/CorpusID:247217035>
- [39] S. Zahara dan Sugianto, "Peramalan Data Indeks Harga Konsumen Berbasis Time Series Multivariate Menggunakan Deep Learning," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 1, hlm. 24–30, Feb 2021, doi: 10.29207/resti.v5i1.2562.
- [40] S. A. Hevit, T. Fahrudin, dan R. Rochmawati, "PERBANDINGAN MODEL ESTIMASI INFLASI PANGAN MENGGUNAKAN ALGORITMA NAIVE FORECASTER DAN EXPONENTIAL SMOOTHING," *Value : Jurnal Manajemen dan Akuntansi*, 2023, [Daring]. Tersedia pada: <https://api.semanticscholar.org/CorpusID:261056510>
- [41] A. L. Haritsah, "Implementasi Model STL (Seasonal Trend Decomposition Procedure based on Loess) dan ARIMA untuk Prediksi Konsentrasi Kualitas Udara," 2015. [Daring]. Tersedia pada: <https://api.semanticscholar.org/CorpusID:158805962>
- [42] N. M. W. Satyawati, I. M. Candiasa, dan N. M. S. Mertasari, "PREDIKSI PENDUDUK MISKIN DI INDONESIA MENGGUNAKAN ANALISIS DEKOMPOSISI," *Delta: Jurnal Ilmiah Pendidikan Matematika*, vol. 9, no. 1, hlm. 77, Jan 2021, doi: 10.31941/delta.v9i1.1248.
- [43] S. Yuni, M. W. Talakua, dan Y. A. Lesnussa, "PERAMALAN JUMLAH PENGUNJUNG PERPUSTAKAAN UNIVERSITAS PATTIMURA AMBON MENGGUNAKAN METODE DEKOMPOSISI," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 9, no. 1, hlm. 41–50, Mar 2015, doi: 10.30598/barekengvol9iss1pp41-50.
- [44] Sunarko, D. B. Syahna, S. Suryanto, dan I. Hamzah, "Decomposition analysis of air temperature data in Bangka Island," dalam *Proceedings of the National Seminar on Nuclear Energy Infrastructure*, dalam Analisis dekomposisi data temperatur udara di Pulau Bangka. Indonesia: Center for Nuclear Energy Systems, National Nuclear Energy Agency, 2019, hlm. 547. [Daring]. Tersedia pada: http://inis.iaea.org/search/search.aspx?orig_q=RN:52064606
- [45] I. K. A. Wirayasa dan H. Santoso, "Analisis Employee Satisfaction Menggunakan Teknik Clustering Dan Classification Machine Learning," *Jurnal Ilmiah Komputer*, vol. 18, no. 1, hlm. 1–10, 2022.
- [46] A. Botchkarev, "Evaluating Performance of Regression Machine Learning Models Using Multiple Error Metrics in Azure Machine Learning Studio," *SSRN Electronic Journal*, 2018, doi: 10.2139/ssrn.3177507.
- [47] D. Chicco, M. J. Warrens, dan G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput Sci*, vol. 7, hlm. e623, Jul 2021, doi: 10.7717/peerj-cs.623.
- [48] A. T. Nurani, A. Setiawan, dan B. Susanto, "Perbandingan Kinerja Regresi Decision Tree dan Regresi Linear Berganda untuk Prediksi BMI pada Dataset Asthma," *Jurnal Sains dan Edukasi Sains*, vol. 6, no. 1, hlm. 34–43, Mei 2023, doi: 10.24246/juses.v6i1p34-43.
- [49] I. Lubis, "ANALISIS MODEL VOLATILITAS INDEKS DAN NILAI MATA UANG DI ASIA TENGGARA," *Jurnal Madani : Ilmu Pengetahuan, Teknologi dan Humaniora*, vol. 1, no. 1, hlm. 123–142, Mar 2018, doi: 10.33753/madani.v1i1.7.
- [50] R. J. Hyndman dan A. B. Koehler, "Another look at measures of forecast accuracy," *Int J Forecast*, vol. 22, no. 4, hlm. 679–688, Okt 2006, doi: 10.1016/j.ijforecast.2006.03.001.
- [51] S. Makridakis, "Accuracy measures: theoretical and practical concerns," *Int J Forecast*, vol. 9, no. 4, hlm. 527–529, Des 1993, doi: 10.1016/0169-2070(93)90079-3.
- [52] S. Kolassa, "Evaluating predictive count data distributions in retail sales forecasting," *Int J Forecast*, vol. 32, no. 3, hlm. 788–803, Jul 2016, doi: 10.1016/j.ijforecast.2015.12.004.



- [53] D. Chicco, M. J. Warrens, dan G. Jurman, “The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation,” *PeerJ Comput Sci*, vol. 7, hlm. e623, Jul 2021, doi: 10.7717/peerj-cs.623.
- [54] P. Mayadewi dan E. Rosely, “PREDIKSI NILAI PROYEK AKHIR MAHASISWA MENGGUNAKAN ALGORITMA KLASIFIKASI DATA MINING,” dalam *Seminar Nasional Sistem Informasi Indonesia*, 2015, hlm. 329–334.
- [55] O. J. Kendek, J. D. Prang, dan M. Paendong, “Prediksi Jumlah Pengunjung Perpustakaan Universitas Sam Ratulangi Manado Menggunakan Metode Dekomposisi,” *d’CARTESIAN*, vol. 3, no. 1, hlm. 73, Mar 2014, doi: 10.35799/dc.3.1.2014.4000.