MATHunesa
Jurnal Ilmiah Matematika *Volume 13 No 01*
e-ISSN : 2716-506X | p-ISSN : 2301-9115 *Tahun 2025*

PEMODELAN CEEMDAN LSTM, CEEMDAN GRU DAN METODE DASARNYA UNTUK PERAMALAN EMISI KARBON DIOKSIDA DI INDONESIA

Eunike Cantika Kusuma Pandi
Program Studi Matematika, FMIPA, Universitas Negeri Semarang
e-mail : enikecantika241@students.unnes.ac.id

Iqbal Kharisudin
Program Studi Matematika, FMIPA, Universitas Negeri Semarang
e-mail : iqbalkharisudin@mail.unnes.ac.id

Abstrak

Emisi karbon dioksida dan dampaknya bagi bumi menjadi fokus penelitian dalam beberapa tahun terakhir. Emisi ini mengalami peningkatan setiap tahun akibat kendaraan bermotor, batu bara, bahan bakar fosil, dan faktor lainnya yang dapat memicu krisis global. Untuk mencapai keseimbangan emisi dan mencegah krisis, diperlukan kompensasi emisi karbon dioksida melalui upaya net zero emissions. Oleh karena itu, penelitian ini bertujuan untuk memprediksi emisi karbon dioksida di masa mendatang menggunakan model peramalan yang tepat.

Penelitian ini mengembangkan model peramalan yang menggabungkan *Long Short Term Memory* (LSTM) dan *Gated Recurrent Unit* (GRU) dengan *Complete Ensemble Mode Decomposition with Adaptive Noise* (CEEMDAN). LSTM adalah pengembangan dari *Recurrent Neural Network* (RNN) yang efektif dalam mengatasi ketergantungan jangka panjang pada data berurutan. GRU, sebagai varian LSTM, menawarkan implementasi yang lebih sederhana dan efisien. CEEMDAN digunakan untuk mengatasi sifat nonlinier dan nonstasioner pada data emisi karbon dioksida, sehingga meningkatkan akurasi model dengan menguraikan sinyal data yang kompleks. Kombinasi model CEEMDAN LSTM, CEEMDAN GRU, dan model dasarnya diterapkan untuk meningkatkan akurasi peramalan.

Hasil penelitian menunjukkan bahwa model EEMD LSTM dan EEMD GRU memberikan hasil peramalan emisi karbon dioksida Indonesia yang lebih baik dan efektif dibandingkan model lainnya. Hal ini dilihat dari nilai *Mean Squared Error* (MSE), *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), dan *Root Mean Square Error* (RMSE) yang paling baik. Hasil ini dapat menjadi dasar pengambilan keputusan untuk kebijakan lingkungan yang lebih baik.
Kata Kunci: *Long Short Term Memory* (LSTM), *Gated Recurrent Unit* (GRU), *Complete Ensemble Mode Decomposition with Adaptive Noise* (CEEMDAN), Emisi Karbon Dioksida.

Abstract

Carbon dioxide emissions and their impact on the earth have been the focus of research in recent years. These emissions are increasing every year due to motor vehicles, coal, fossil fuels, and other factors that can trigger a global crisis. To achieve emission balance and prevent a crisis, carbon dioxide emission compensation through net zero emissions is needed. Therefore, this study aims to predict future carbon dioxide emissions using an appropriate forecasting model.

This research develops a forecasting model that combines Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) with Complete Ensemble Mode Decomposition with Adaptive Noise (CEEMDAN). LSTM is a development of Recurrent Neural Network (RNN) that is effective in overcoming long-term dependency on sequential data. GRU, as a variant of LSTM, offers a simpler and more efficient implementation. CEEMDAN is used to address the nonlinear and nonstationary nature of carbon dioxide emission data, thereby improving model accuracy by decomposing complex data signals. A combination of CEEMDAN LSTM, CEEMDAN GRU, and their base models is applied to improve forecasting accuracy.

The results show that EEMD LSTM and EEMD GRU models provide better and more effective forecasting results for Indonesia's carbon dioxide emissions than other models. This is seen from the best Mean Squared Error (MSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE) values. These results can be the basis for decision making for better environmental policies.
Keywords: *Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU), Complete Ensemble Mode Decomposition with Adaptive Noise (CEEMDAN), Carbon Dioxide Emissions.*

PENDAHULUAN

Di tengah zaman yang terus berkembang, keputusan untuk masa depan sering dihadapkan dengan risiko dan ketidakpastian. Untuk mengatasinya, berbagai model pendekatan digunakan, salah satunya adalah data mining yang mampu menganalisis data dalam jumlah besar untuk menemukan pola serupa dengan data asli. Salah satu metode data mining yang populer adalah peramalan yang digunakan untuk memprediksi kejadian di masa depan. Dalam peramalan, metode time series sangat umum digunakan, terutama dengan algoritma *deep learning* seperti *Recurret Neural Network* (RNN). Namun, RNN memiliki kelemahan dalam mengingat data jangka panjang, sehingga dikembangkan model *Long Short Term Memory* (LSTM) dan *Gated Recurrent Unit* (GRU) untuk mengatasi masalah tersebut.

LSTM adalah model yang dirancang untuk mengatasi masalah gradien yang hilang dalam RNN, dengan kemampuan untuk mengingat informasi dalam jangka panjang. LSTM memproses dan menangani input dan ouput yang panjang pada tiap waktu (Setiawan & Kartikasari, 2022). Namun, LSTM memiliki kelemahan dalam hal kompleksitas dan waktu pelatihan. Untuk mengatasi ini, dikembangkan model GRU yang lebih sederhana dan efisien (Ilham Caniago et al., 2021). Untuk memperbaiki prediksi data nonlinier dan nonstasioner, LSTM dan GRU sering digabungkan dengan metode CEEMDAN yang membantu menguraikan data tersebut. Salah satu data yang memiliki sifat tersebut dan sedang menjadi perbincangan adalah emisi karbon dioksida.

Emisi karbon dioksida menjadi perhatian global karena dampaknya terhadap perubahan iklim dan kesehatan manusia. Banyak negara, termasuk Indonesia, berkomitmen untuk mencapai *net zero emissions* untuk menyeimbangkan emisi karbon dioksida. Indonesia menargetkan pengurangan emisi sebesar 31,89 pada tahun 2030 dan *net zero emissions* pada tahun 2060 (Putri, 2022). Peningkatan emisi ini disebabkan oleh berbagai faktor, termasuk peningkatan jumlah kendaraan. Upaya untuk mengurangi emisi ini memerlukan langkah konkret dari pembuat kebijakan di seluruh dunia, dengan memanfaatkan data prediksi emisi untuk merencanakan tindakan preventif.

Dalam penelitian ini, akan dilakukan peramalan emisi karbon dioksida di Indonesia menggunakan model CEEMDAN LSTM dan CEEMDAN GRU. Kedua metode ini akan dibandingkan untuk menentukan model yang paling akurat dalam memprediksi emisi karbon dioksida di Indonesia. Diharapkan, hasil penelitian ini dapat memberikan kontribusi dalam upaya pengurangan emisi karbon dioksida di Indonesia dan mencapai target *net zero emissions* yang telah ditetapkan.

KAJIAN TEORI

NEAREST NEIGHBOR

Nearest Neighbor merupakan Imputasi tetangga terdekat univariat mungkin merupakan skema yang paling sederhana yang tersedia, karena titik akhir dari kesenjangan digunakan sebagai estimasi untuk semua nilai yang hilang (Junninen et al., 2004).

$$y = y_1 \text{ jika } x \leq x_1 + \frac{(x_2 - x_1)}{2} \quad (1)$$

$$y = y_2 \text{ jika } x \leq x_1 + \frac{(x_2 - x_1)}{2}$$

dimana y adalah interpolasi, x adalah titik waktu dari interpolasiy_1 dan x_1 adalah koordinat titik awal celah, dan y_2 dan x_2 adalah koordinat titik akhir celah.

INTERQUARTILE RANGE

Interquartile Range (IQR) merupakan metode yang menghitung rentang interkuartil dari pengamatan kekuatan sinyal terima (RSS) deret waktu, yaitu selisih antara persentil ke-75 dan persentil ke-25. IQR membagi data menjadi kuartil Q1 sebagai persentil ke- 25, Q2 sebagai median atau persentil ke- 50, dan Q3 sebagai persentil ke- 75. Berikut persamaan IQR untuk mendeteksi outlier (Takahashi et al., 2024; Yaro et al., 2023).

$$IQR = Q_3 - Q_1 \quad (2)$$

Hitung Q1, Q3, dan IQR. Selanjutnya identifikasi batas bawah dan batas atas batas bawah .

$$\tau_{IQR,lower} = Q_1 - 1{,}5 \times IQR \quad (3)$$

$$\tau_{IQR,upper} = Q_3 + 1{,}5 \times IQR \quad (4)$$

Pengamatan RSS di luar rentang IQR dianggap senagai outlier. Ambang batas ditetapkan di luar 1,5 kali IQR dari rentang Q1 dan Q3, yang mencakup 99,3 data pada distribusi normal, dan ini juga digunakan dalam boxplot.

NORMALISASI *MIN-MAX*

Teknik normalisasi *min-max* adalah metode populer dalam analisis regresi yang mengubah data dalam rentang $[0,1]$ atau $[-1,1]$ menggunakan nilai minimum dan maksimum dari kumpulan data (Kammoun et al., 2024; Tawakuli et al., 2024). Jika diberikan deretan data $X = [x_1, x_2, x_3, \dots, x_n]$, data dinormalisasi dengan rentang dihitung dengan rentang $[0,1]$ dan n sebagai panjang data, berikut persamaannya (Mohanty & Dash, 2023).

$$x_{norm_i} = \frac{x_i - \min(X)}{\max(X) - \min(X)}, i = 1, \dots, n \quad (5)$$

Metode ini optimal untuk data yang sedikit atau tanpa outlier dan terdistribusi merata di seluruh rentang minimum dan maksimum. Normalisasi dapat disesuaikan dalam rentang $[a, b]$ dengan persamaan yang disesuaikan.

$$x_{norm_i} = \frac{x_i - \min(X)}{\max(X) - \min(X)}, (b - a) \quad (6)$$

Dalam penelitian ini, metode *min-max* digunakan untuk menormalkan data emisi karbon dioksida dalam rentang $[0, 1]$. Setelah normalisasi, diperlukan proses denormalisasi untuk mengembalikan data ke skala aslinya, yang merupakan kebalikan dari proses normalisasi.

COMPLETE ENSEMBLE EMPIRICAL MODE DECOMPOSITIONWITH ADAPTIVE NOISE

Model *Complete Ensemble Empirical Mode corruption with Adaptive Noise* (CEEMDAN) adalah pengembangan dari metode *Empirical Mode corruption* (EMD) dan *Ensemble Empirical Mode Decomposition* (EEMD). EMD awalnya diusulkan oleh Huang et al. (1998) untuk menguraikan deret nonlinear menjadi fungsi modus intrinsik (IMF) dan residu yang memiliki masalah pencampuran mode. Hal ini diatasi oleh EEMD dengan menambahkan *white noise Gaussian* namun memiliki kekurangan seperti jumlah mode yang bervariasi.

CEEMDAN diusulkan untuk memperbaiki masalah EEMD dengan memberikan pemisahan mode yang lebih baik dan sinyal rekonstruksi yang lebih akurat dengan biaya komputasi lebih rendah. Teknik ini melibatkan penambahan white noise adaptif pada setiap tingkat dekomposisi sinyal, mengurangi pencampuran mode dan memastikan efek dekomposisi sinyal yang akurat. Langkah-langkah CEEMDAN sebagai berikut.

$$X^i(t) = X(t) + X_0 N^i(t), i = 1, 2, \dots, n \quad (7)$$

Menambahkan *white noise* $N^i(t)$ pada data asli untuk membentuk sinyal baru. $X(t)$ sebagai tinggi gelombang signifikan, t sebagai titik waktu, i sebagai *white noise* ke i, X_0 sebagai deviasi *noise* standar, dan $X^i(t)$ sebagai sinyal baru yang dihasilkan.

$$IMF_1(t) = \frac{1}{n}\sum_{i=1}^{n} IMF_1^i(t) \quad (8)$$

$$r_1(t) = X(t) - IMF_1(t) \quad (9)$$

Menguraikan sinyal ini untuk mendapatkan IMF pertama dan residu $r_1(t)$.

$$r_1^i(t) = r_1(t) + N^i(t) \quad (10)$$

$$IMF_2(t) = \frac{1}{n}\sum_{i=1}^{n} IMF_2^i(t) \quad (11)$$

$$r_2(t) = r_1(t) - IMF_2(t) \quad (12)$$

Menambahkan white noise pada residu dan menguraikannya untuk mendapatkan IMF kedua.

$$IMF_k(t) = \frac{1}{n}\sum_{i=1}^{n} IMF_k^i(t) \quad (13)$$

$$r_k(t) = r_k(t) - IMF_k(t) \quad (14)$$

$$X(t) = \sum_{i=1}^{k} IMF_n(t) + R(t) \quad (15)$$

Melanjutkan proses ini hingga residu tidak dapat didekomposisi lebih lanjut. Proses ini menghasilkan IMF yang terpisah dengan baik dan residu akhir yang merupakan fungsi monotonik.

LONG SHORT TERM MEMORY

Long Short Term Memory (LSTM) adalah model *deep learning* yang didasarkan pada *Recurret Neural Network* (RNN) dengan tambahan elemen memori untuk menangani ketergantungan jangka panjang dan mengatasi masalah gradien yang hilang . Menurut Kumari & Singh (2022) LSTM terdiri dari *cell, input gate, forget gate*, dan *output gate* yang mengatur aliran informasi dalam model. *Cell* menyimpan nilai untuk jangka waktu tertentu, sementara *input gate* mengontrol masuknya informasi baru, *forget gate* mengontrol seberapa banyak informasi yang disimpan, dan *output gate* mengontrol informasi yang keluar dari *cell*. Model ini

unggul dalam memproses, mengklasifikasikan, dan meramalkan data deret waktu, mengatasi masalah yang ada pada RNN tradisional, seperti ledakan dan penurunan gradien. Berikut persamaan-persamaan model LSTM.

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (16)$$

f_t merupakan perhitungan nilai *input gate*. Dua persamaan dibawah ini digunakan pada sel memori dalam menggunakan fungsi sigmoid untuk memperbarui data dan menyimpan nilai baru dengan fungsi aktivasi tahunan.

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i \quad (17)$$

$$c_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (18)$$

$$c_t = f_t * c_{t-1} + i_t * c_t \quad (19)$$

c_t merupakan kombinasi *input gate. Output gate* memiliki dua *gate* untuk menentukan nilai yang ditampilkan dengan menghitung dua persamaan terakhir berikut.

$$O_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (20)$$

$$h_t = O_t . tanh(c_t) \quad (21)$$

GATED RECURRENT UNIT

Gated Recurrent Unit (GRU) merupakan model perbaikan dari RNN yang dirancang untuk menangani masalah penerjemahan mesin dan telah diterapkan dalam berbagai bidang (Cho et al., 2014; Lv et al., 2023a). GRU mengatasi masalah ledakan dan hilangnya gradien yang terjadi pada RNN dan menyederhanakan perhitungan kompleks yang ada pada LSTM dengan menggabungkan *input gate* dan *forget gate* menjadi satu *update gate*. GRU juga memiliki *reset gate* yang menentukan apakah keadaan tersembunyi dari masa lalu perlu diabaikan. Keunggulan utama GRU adalah struktur yang lebih sederhana dan kecepatan komputasi yang lebih cepat dibandingkan dengan LSTM, sambil tetap mempertahankan efektivitas dalam memori selektif informasi jangka panjang dan pendek. r_t pada *update gate* dan z_t pada *reset gate* diperoleh dengan persamaan berikut (Lv et al., 2023a).

$$\sigma_{(x)} = \frac{1}{1+e^x} \quad (22)$$

$$R_t = \sigma(W_r x_t + U_r h_{t-1}) \quad (23)$$

$$Z_t = \sigma(W_z x_t + U_z h_{t-1}) \quad (24)$$

U dan W adalah parameter bobot (Lv et al., 2023b). h_t sebagai keadaan tersembunyi, $\widetilde{h_t}$ sebagai himpunan kadidat, dan $Tanh$ sebagai fungsi garis singgung hiperbolik persamaannya adalah sebagai berikut.

$$H_t = (1 - z_t)h_{t-1} + z_t\widetilde{h_t} \quad (25)$$

$$\widetilde{H_t} = \tanh(W_h x_t + U_h(r_t \odot h_{t-1})) \quad (26)$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (27)$$

CEEMDAN LSTM

Penelitian ini mengusulkan model kombinasi CEEMDAN dan LSTM untuk memprediksi emisi karbon dioksida yang memiliki karakteristik nonlinier, nonstasioner, dan heterogenitas waktu-frekuensi. Model ini memanfaatkan keunggulan CEEMDAN untuk menyesuaikan sinyal emisi karbon yang tidak mulus dan tidak beraturan dengan menguraikannya menjadi beberapa mode IMF yang berbeda (Yun et al., 2022). Kemudian LSTM digunakan untuk meningkatkan kemampuan peramalan terhadap informasi IMF yang diperoleh. Proses prediksi ini melibatkan tiga tahap. Pertama ada dekomposisi CEEMDAN digunakan untuk menguraikan emisi karbon asli menjadi sinyal modus IMF dan residu, yang mencerminkan volatilitas frekuensi waktu emisi karbon. Kedua, prediksi LSTM diterapkan untuk menyesuaikan tren nonlinier dari sinyal modus dan residu, mengungkapkan kinerja prediksi setiap sinyal skala waktu. Terakhir adanya integrasi Nilai prediksi akhir dihitung dengan menjumlahkan hasil prediksi dari setiap mode IMF dan residu.

CEEMDAN GRU

Penelitian ini mengusulkan model CEEMDAN GRU untuk meningkatkan akurasi prediksi emisi karbon dioksida. Model ini mengikuti proses "dekomposisi – prediksi - rekonstruksi". CEEMDAN digunakan untuk menguraikan data emisi menjadi beberapa urutan dengan frekuensi terbatas dari tinggi ke rendah, mencakup karakteristik lokal dan sisa dari data asli (Zhang & Yang, 2020). Setelah itu, GRU digunakan untuk memprediksi urutan frekuensi tersebut.

MEAN SQUARED ERROR

Mean Squared Error (MSE) merupakan metode dalam analisis evaluasi yang menghitung kesalahan kuadrat rata-rata. Pertama-tama hitung kuadrat kesalahan residual setiap titik data dan kemudian menghitung rata-ratanya, berikut persamaannya.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (28)$$

$(y - \hat{y})$ merupakan kesalahan residual dengan y dan $\hat{y}$ menunjukkan nilai aktual dan prediksi. Bakay & Ağbulut (2021) menunjukkan bahwa nilai MSE dapat bervariasi dari 0 hingga ∞ di mana nilai yang lebih kecil lebih disukai.

MEAN ABSOLUTE PERCENTAGE ERROR

Mean Absolute Percentage Error (MAPE) adalah metrik yang menambahkan persentase pada *Mean Absolute Error* (MAE) (Kumar et al., 2020). Model yang baik memiliki nilai MAPE lebih rendah dibandingkan model lainnya. Berikut persamaan dari MAPE.

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (29)$$

$MAPE < 10\%$: akurasi prediksi tinggi.
$MAPE\ 10\% - 20\%$: akurasi prediksi baik.
$MAPE\ 20\% - 50\%$: akurasi prediksi wajar.
$MAPE > 50\%$: akurasi prediksi tidak akurat

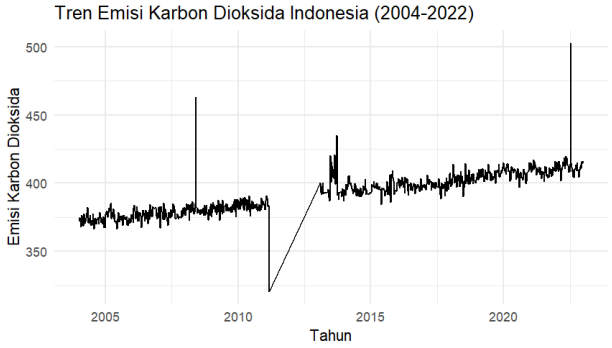
METODE

Penelitian ini menggunakan pendekatan kuantitatif deskriptif untuk menganalisis dan memprediksi emisi karbon dioksida di Indonesia. Data yang digunakan adalah data sekunder emisi karbon dioksida mingguan di Indonesia, dari tahun 2004 hingga 2022 dengan total 855 data. Data ini diperoleh dari *National Oceanic and Atmospheric Administration* (NOAA). Data diakses melalui situs resmi (https://www.noaa.gov/).

Untuk memprediksi emisi karbon dioksida dengan metode CEEMDAN LSTM, CEEMDAN GRU dan metode dasarnya, dilakukan langkah- langkah berikut.

1. Pengumpulan data sekunder diperoleh dari sumber resmi NOAA. Data ini mencakup emisi karbon dioksida mingguan selama periode penelitian.
2. Preprocessing data meliputi penanganan nilai yang hilang, penghilangan outlier, dan normalisasi data untuk menjaga konsistensi.
3. Dekomposisi EMD, EEMD, CEEMDAN untuk menguraikan data emisi karbon sehingga mendapatkan sinyal modus IMF dan residu.
4. Pembagian data menjadi dua set, yaitu data t*raining* sebesar 80% dan data *testing* sebesar 20% untuk pelatihan dan pengujian model.
5. Pemodelan dengan LSTM dan GRU diterapkan untuk memprediksi emisi karbon dioksida berdasarkan sinyal yang telah diolah.
6. Pengukuran akurasi dan performa model dievaluasi menggunakan metrik MSE.
7. Peramalan emisi karbon dioksida dengan model terbaik untuk meramalkan emisi karbon dioksida lima tahun ke depan dan tren emisi masa depan.

HASIL DAN PEMBAHASAN



Gambar 1. Plot Data Asli

Pada Gambar 1, data emisi karbon dioksida yang diterima memiliki beberapa data minggu yang hilang dan terdapat beberapa *outlier*. Untuk itu diperlukan *preprocessing* data untuk membuat data lebih lengkap dan efisien.

Langkah pertama *preprocessing* data adalah mendeteksi dan mengatasi ketidaklengkapan data mingguan sehingga setiap bulan memiliki setidaknya empat data mingguan. Setelah dilakukan input data tanggal dan nilai emisi yang hilang menggunakan *Nearest Neighbor*, data yang awalnya sebanyak 855 sekarang bertambah menjadi 993.

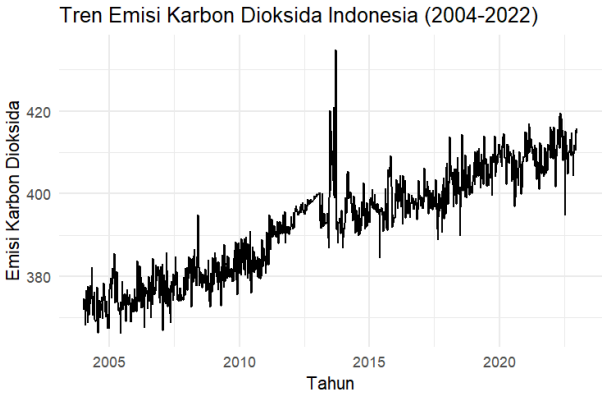
Selanjutnya dilakukan deteksi *outlier* data menggunakan metode IQR. Pada Tabel 1, terlihat bahwa terdapat tiga *outlier* yang tersebar acak pada data.

Tabel 1. List *Outlier*

No.	Tanggal	Emisi (ppm)
1.	2008-05-27	462.87
2.	2011-03-01	320.67
3.	2022-07-12	502.60

Tiga outlier tersebut diganti nilainya dengan nilai median data yaitu 394.875 ppm. Setelah pengantian *outlier*, data emisi karbon dioksida terendah berada

pada tanggal 27 Juli 2004 sebanyak 336.31 ppm dan data emisi karbon dioksida tertinggi berada pada tanggal 17 September 2013 sebanyak 434.91 ppm. Gambar 2 menampilan data emisi karbon dioksida setelah dilakukan penanganan *missing value* dan *outlier*.



Gambar 2. Plot Data Tanpa *Missing Value* dan *Outlier*

Sebelum melanjutkan proses pemodelan, data akan dinormalisasi menggunakan metode min-max. Metode normalisasi *min-max* ini menggunakan rentang [0,1]. Normalisasi memudahkan proses pemodelan selanjutnya.

Langkah pertama dalam analisis peramalan ini adalah menggunakan metode CEEMDAN dan metode dasarnya yaitu EMD dan EEMD. Untuk mendapatkan hasil yang optimal, diperlukan parameter yang tepat dalam dekomposisi EMD, EEMD, dan CEEMDAN. Parameter yang akan dicari meliputi *ensemble size* dan *noise strength* dengan nilai MSE paling rendah. Pada Tabel 2 ditunjukan bahwa *ensemble size* 200 dan *noise strength* 0.2 merupakan nilai parameter yang paling baik dalam melakukan dekomposisi. Nilai dari parameter-parameter tersebut digunakan dalam dekomposisi EMD, EEMD, dan CEEMDAN.

Tabel 2. MSE untuk Setiap Parameter CEEMDAN

Ensember Size	*Noise Strength*	MSE
100	0.1	3.260513×10^{-33}
100	0.2	3.977947×10^{-33}
100	0.4	3.170808×10^{-33}
200	0.1	3.667788×10^{-33}
200	**0.2**	$\mathbf{2.808394 \times 10^{-33}}$
200	0.4	3.614574×10^{-33}

Data emisi karbon dioksida selanjutnya dibagi terlebih dahulu sebelum dimodelkan LSTM atau GRU. Pembagian data dibagi menjadi 80% data *training* dan 20% data *testing*, dengan 794 data untuk *training* dan 199 data untuk *testing*.

Dalam memodelkan data menggunakan LSTM dan GRU, perlu dilakukan persiapan dan penyesuaian data. Langkah ini melibatkan pembuatan dataset sebagai input dan dan *time steps* sebagai panjang urutan input yang diproses oleh model dalam satu waktu selama pelatihan. Penelitian ini menggunakan *time steps* bernilai 10 yang berarti model akan melihat 10 langkah waktu berturut-turut untuk membuat prediksi. Persiapan ini memungkinkan model LSTM atau GRU untuk mempelajari pola dalam urutan data dan menangkap pola jangka pendek ataupun jangka panjang

Seperti halnya model CEEMDAN, model LSTM dan GRU memerlukan parameter terbaik untuk memaksimalkan kinerja. Parameter penting termasuk *unit* dan *epoch*. Percobaan dilakukan dengan mengiterasi *unit* dan *epoch* untuk menemukan konfigurasi yang optimal.

Tabel 2. MSE untuk Setiap Parameter CEEMDAN LSTM dan CEEMDAN GRU

	LSTM	GRU
Unit 32, *epoch* 50	0.01349049	0.01343585
Unit 32, *epoch* 100	0.01346387	0.01343776
Unit 32, *epoch* 200	0.01342702	0.01342512
Unit 64, *epoch* 50	0.01346061	0.01344149
Unit 64, *epoch* 100	0.01344671	0.01343674
Unit 64, *epoch* 200	0.01342371	0.01342767
Unit 128, *epoch* 50	0.01345356	0.01343889
Unit 128, *epoch* 100	0.01343831	0.01343638
Unit 128, epoch 200	**0.01342332**	**0.01342734**

Berdasarkan Tabel 3, parameter terbaik untuk model CEEMDAN LSTM dan CEEMDAN GRU adalah *unit* 128 dan *epoch* 200.

Setelah menentukan parameter terbaik, langkah selanjutnya adalah menerapkannya pada model LSTM dan GRU. Proses ini dimulai dengan pembuatan fungsi pemodelan, diikuti dengan pembangunan dan pelatihan model, dan diakhiri dengan menjalankan model untuk menghasilkan peramalan. Setelah pemodelan, diperlukan proses denormalisasi untuk mengembalikan nilai data ke bentuk awal sebelum dilakukan evaluasi model. Hasil peramalan model CEEMDAN LSTM dan CEEMDAN GRU serta metode dasarnya setelah dilakukan denormalisasi dapat di lihat di Tabel 3 dan Tabel 4.

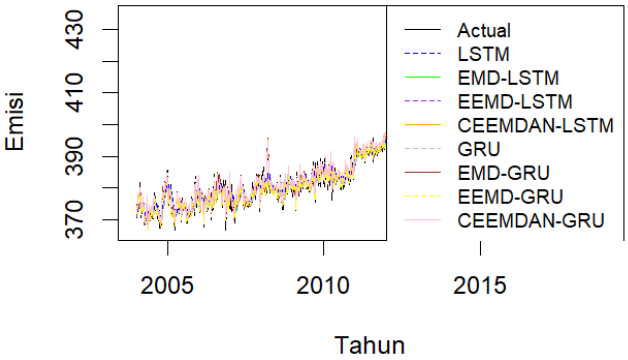
Tabel 4. Tabel Perbandingan Hasil Peramalan Data *Training*

Tanggal	2004-01-08	2004-01-14	2004-01-21	2004-01-29	2004-02-04
Data Asli	371.77	374.63	374.31	368.20	373.69
LSTM	374.43	372.50	372.72	372.95	373.46
EMD LSTM	373.63	373.11	374.87	374.33	374.23
EEMD LSTM	371.88	371.10	373.16	376.20	377.21
CEEMDAN LSTM	371.88	372.62	374.53	376.45	377.75
GRU	373.22	372.20	372.00	372.23	372.71
EMD GRU	374.11	374.21	375.56	374.57	374.38
EEMD GRU	372.28	371.87	374.83	376.13	377.00
CEEMDAN GRU	372.58	372.77	375.43	377.13	378.68

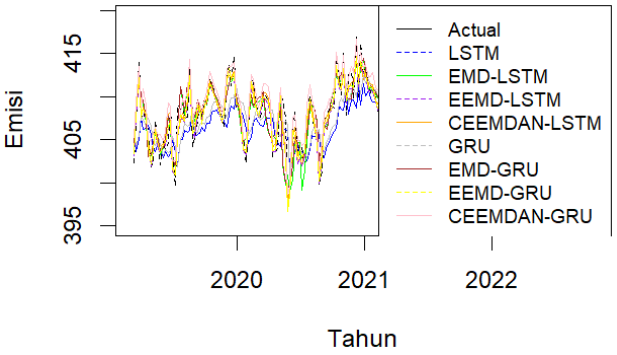
Tabel 5. Tabel Perbandingan Hasil Peramalan Data *Testing*

Tanggal	2019-03-12	2019-03-19	2019-03-26	2019-04-02	2019-04-09
Data Asli	404.19	410.91	407.77	406.12	401.49
LSTM	405.79	403. 64	405.10	407.64	406.19
EMD LSTM	404.49	407.93	411.72	410.22	409.30
EEMD LSTM	403.14	407.87	412.04	407.67	408.69
CEEMDAN LSTM	404.78	408.83	412.10	407.62	408.85
GRU	406.22	404.21	405.53	407.69	407.34
EMD GRU	404.36	408.24	412.05	410.15	409.61
EEMD GRU	403.42	408.07	412.53	407.63	408.65
CEEMDAN GRU	405.28	409.31	413.56	408.74	410.88

Hasil permalan tiap-tiap model selanjutnya diplotkan. Berikut grafik dari hasil *plotting* data peramalan baik data *testing* maupun data *training* emisi karbon dioksida.



Gambar 3. Plot Perbandingan Permalan untuk Data *Training*



Gambar 4. Plot Perbandingan Permalan untuk Data *Testing*

Dari Gambar 3 dan Gambar 4, dapat dilihat bahwa plot dari seluruh model memiliki kemiripan dengan plot data asli. Namun model EEMD LSTM dan EEMD GRU memiliki kemiripan paling tinggi dengan data asli.

Selanjutnya model CEEMDAN LSTM dan CEEMDAN GRU, beserta metode dasarnya, dianalisis menggunakan MSE dan MAPE. Analisis ini dilakukan untuk menentukan model mana yang paling baik dalam meramalkan data emisi karbon dioksida. Hasil analisi MSE dan MAPE tiap model sebagai berikut.

Tabel 6. Hasil Analisis Data *Training*

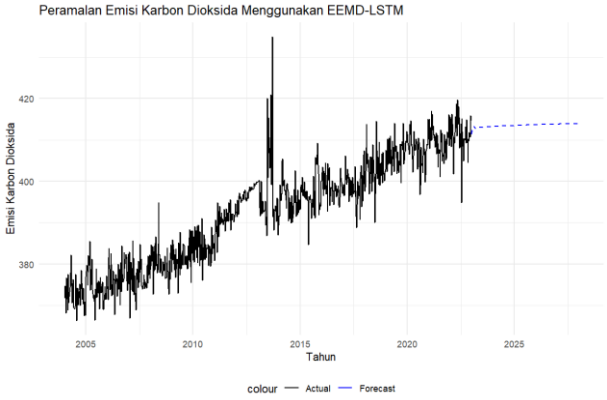
Model	MSE	MAPE
LSTM	15.744444	0.007109065
EMD LSTM	5.136561	0.004538709
EEMD LSTM	**2.515999**	**0.002972399**
CEEMDAN LSTM	2.745743	0.003278550
GRU	15.095471	0.007050694
EMD GRU	5.273784	0.004624249
EEMD GRU	2.563498	0.003006650
CEEMDAN GRU	4.081534	0.004247637

Tabel 7. Hasil Analisis Data *Testing*

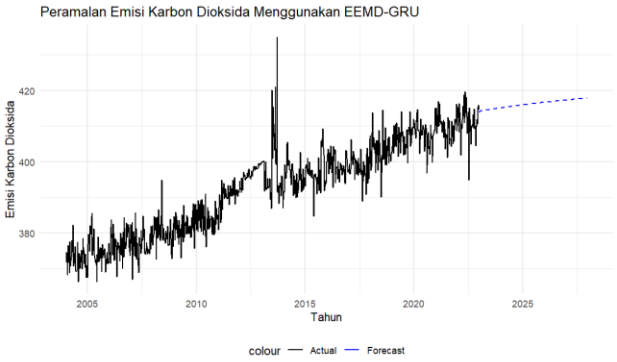
Model	MSE	MAPE
LSTM	16.436126	0.007774739
EMD LSTM	5.457740	0.004569706
EEMD LSTM	2.850140	0.002971014
CEEMDAN LSTM	2.974044	0.003094562
GRU	14.402208	0.007072077
EMD GRU	5.118413	0.004443875
EEMD GRU	**2.537033**	**0.002801445**
CEEMDAN GRU	4.581640	0.003844191

Berdasarkan hasil analisis di Tabel 6, model EEMD LSTM memiliki nilai evaluasi terbaik dengan MSE 2.515999 dan MAPE 0.002972399, yang merupakan nilai terendah dibandingkan model lainnya. Sama halnya dalam analisis data testing pada Tabel 7, model EEMD GRU menunjukkan performa terbaik dalam meramalkan emisi karbon dioksida, dengan MSE 2.537033 dan MAPE 0.002801445. Penambahan metode CEEMDAN pada model LSTM dan GRU menyebabkan *overfitting*, sehingga tidak memberikan dampak signifikan dan menurunkan kinerja model gabungan EEMD (Bouhalais et al., 2018). Dengan mempertimbangkan penghilangan CEEMDAN, model gabungan EEMD dengan model LSTM dan GRU memberikan hasil peramalan yang lebih baik.

Model EEMD LSTM dan EEMD GRU digunakan untuk meramalkan data emisi karbon dioksida selama lima tahun ke depan. Selain algoritma pemodelan LSTM dan GRU yang telah dibahas di bagian sebelumnya, akan dilakukan juga pemodelan khusus untuk meramalkan data lima tahun ke depan. Berikut grafik hasil *plotting* data permalan lima tahun oleh model EEMD LSTM dan EEMD GRU.



Gambar 5. Plot Perbandingan Permalan Lima Tahun Kedepan EEMD LSTM



Gambar 6. Plot Perbandingan Permalan Lima Tahun Kedepan EEMD GRU

Berdasarkan Gambar 5 dan Gambar 6, plot model EEMD LSTM dan EEMD GRU menunjukkan bahwa akan ada kenaikan emisi karbon dioksida di Indonesia dalam lima tahun mendatang.

PENUTUP

SIMPULAN

Berdasarkan hasil penelitian di pembahasan, didapat bahwa model CEEMDAN optimal untuk meramalkan emisi karbon dioksida menggunakan parameter tertentu seperti *ensemble size* 200 dan *noise strength* 0.2. Model CEEMDAN LSTM dan CEEMDAN GRU paling efektif menggunakan *unit* 128 dan *epoch* 200. Berdasarkan evaluasi MSE dan MAPE, model EEMD LSTM dan EEMD GRU terbukti lebih unggul dibandingkan kombinasi model CEEMDAN LSTM, CEEMDAN GRU, dan model dasar lainnya. Prediksi untuk 5 tahun ke depan menunjukkan peningkatan emisi dengan model EEMD LSTM dan EEMD GRU.

SARAN

Berdasarkan kesimpulan, disarankan agar model LSTM dan GRU digunakan dengan data harian emisi karbon dioksida untuk hasil yang lebih optimal. Penelitian mendatang sebaiknya menggunakan script dalam bahasa *Python* karena lebih mendukung paket yang relevan. Selain itu, pengujian dan penyetelan parameter seperti jumlah lapisan, laju pembelajaran, ukuran *batch*, dan tingkat *dropout* perlu dilakukan untuk mengurangi risiko *overfitting* dan mengoptimalkan model. Pemerintah Indonesia juga diharapkan melakukan upaya untuk menekan emisi karbon dioksida guna mencapai program *net zero*

emissions pada tahun 2060 dan pengurangan emisi sebesar 31,89% pada tahun 2030.

DAFTAR PUSTAKA

Bakay, M. S., & Ağbulut, Ü. (2021). Electricity production based forecasting of greenhouse gas emissions in Turkey with deep learning, support vector machine and artificial neural network algorithms. *Journal of Cleaner Production*, *285*, 125324. https://doi.org/10.1016/j.jclepro.2020.125324

Bouhalais, M. L., Djebala, A., Ouelaa, N., & Babouri, M. K. (2018). CEEMDAN and OWMRA as a hybrid method for rolling bearing fault diagnosis under variable speed. *International Journal of Advanced Manufacturing Technology*, *94*(5–8), 2475–2489. https://doi.org/10.1007/s00170-017-1044-0

Cho, K., van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder–decoder approaches. *Proceedings of SSST 2014 - 8th Workshop on Syntax, Semantics and Structure in Statistical Translation*, 103–111. https://doi.org/10.3115/v1/w14-4012

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Snin, H. H., Zheng, Q., Yen, N. C., Tung, C. C., & Liu, H. H. (1998). The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *454*(1971), 903–995. https://doi.org/10.1098/rspa.1998.0193

Ilham Caniago, A., Kaswidjanti, W., & Pembangunan Nasional Veteran Yogyakarta, U. (2021). Recurrent Neural Network dengan Gate Recurrent Unit Untuk Prediksi Harga Saham. *Jurnal Informatika Dan Teknologi Informasi*, *18*(3), 345–360. https://doi.org/10.31515/telematika.v18i3.6650

Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J., & Kolehmainen, M. (2004). Methods for imputation of missing values in air quality data sets. *Atmospheric Environment*, *38*(18), 2895–2907. https://doi.org/10.1016/j.atmosenv.2004.02.026

Kammoun, A., Ravier, P., & Buttelli, O. (2024). Impact of PCA Pre-Normalization Methods on Ground Reaction Force Estimation Accuracy. *Sensors*, *24*(4). https://doi.org/10.3390/s24041137

Kumar, R., Kumar, P., & Kumar, Y. (2020). Time Series Data Prediction using IoT and Machine Learning Technique. *Procedia Computer Science*, *167*(2019), 373–381. https://doi.org/10.1016/j.procs.2020.03.240

Kumari, S., & Singh, S. K. (2022). Machine learning-based time series models for effective CO2 emission prediction in India. *Environmental Science and Pollution Research*, *0123456789*. https://doi.org/10.1007/s11356-022-21723-8

Lv, Z., Wang, N., Lou, R., Tian, Y., & Guizani, M. (2023a). Towards carbon Neutrality: Prediction of wave energy based on improved GRU in Maritime transportation. *Applied Energy*, *331*(September 2022), 120394. https://doi.org/10.1016/j.apenergy.2022.120394

Lv, Z., Wang, N., Lou, R., Tian, Y., & Guizani, M. (2023b). Towards carbon Neutrality: Prediction of wave energy based on improved GRU in Maritime transportation. *Applied Energy*, *331*(September 2022), 120394. https://doi.org/10.1016/j.apenergy.2022.120394

Mohanty, S., & Dash, R. (2023). A New Dual Normalization for Enhancing the Bitcoin Pricing Capability of an Optimized Low Complexity Neural Net with TOPSIS Evaluation. *Mathematics*, *11*(5). https://doi.org/10.3390/math11051134

Putri, C. (2022). *Target Baru! RI Bakal Kurangi Emisi Karbon Jadi 31,9% di 2030*. CNBC Indonesia. https://www.cnbcindonesia.com/news/20221026154840-4-382700/target-baru-ri-bakal-kurangi-emisi-karbon-jadi-319-di-2030

Setiawan, Y., & Kartikasari, P. (2022). *Prediksi Harga Jual Kakao dengan Metode Long Short-Term Memory Menggunakan Metode Optimasi Root Mean Square Propagation dan Adaptive Moment Estimation Dilengkapi Gui Rshiny*. *11*(1), 99–107. https://ejournal3.undip.ac.id/index.php/gaussian/

Takahashi, K., Ooka, R., & Kurosaki, A. (2024). Seasonal threshold to reduce false positives for prediction-based outlier detection in building energy data. *Journal of Building Engineering*, *84*. https://doi.org/10.1016/j.jobe.2024.108539

Tawakuli, A., Havers, B., Gulisano, V., Kaiser, D., & Engel, T. (2024). Survey:Time-series data preprocessing: A survey and an empirical analysis. *Journal of Engineering Research (Kuwait)*. https://doi.org/10.1016/j.jer.2024.02.018

Yaro, A. S., Maly, F., & Prazak, P. (2023). Outlier Detection in Time-Series Receive Signal Strength Observation Using Z-Score Method with Sn Scale Estimator for Indoor Localization. *Applied Sciences (Switzerland)*, *13*(6). https://doi.org/10.3390/app13063900

Yun, P., Huang, X., Wu, Y., & Yang, X. (2022). Forecasting carbon dioxide emission price using a novel mode decomposition machine learning hybrid model of CEEMDAN-LSTM. *Energy Science & Engineering*, *11*(1), 79–96. https://onlinelibrary.wiley.com/doi/full/10.1002/ese3.1304

Zhang, X., & Yang, Y. (2020). Suspended sediment concentration forecast based on CEEMDAN-GRU model. *Water Science and Technology: Water Supply*, *20*(5), 1787–1798. https://doi.org/10.2166/ws.2020.087