

# Peramalan Emisi Karbon Menggunakan Metode SARIMA dan LSTM

(Forecasting Carbon Emissions Using the SARIMA and LSTM Methods)

Syifa Ilma Nabila Suwandi\*, Raras Tyasnurita, Hanifan Muhayat

Dept Sistem Informasi, Institut Teknologi Sepuluh Nopember

Jl. Teknik Kimia, Keputih, Kec. Sukolilo, Kota SBY, Jawa Timur, INDONESIA

Email: syifasuwandi.18052@mhs.its.ac.id, raras@is.its.ac.id, hanifanmuhayat.18052@mhs.its.ac.id

\*Penulis Korespondensi

**Abstract** The majority of greenhouse gas (GHG) effects are caused by very high levels of carbon emissions in the world. Therefore, it is necessary to take action to control the levels of carbon emissions in the world. In this study, the world's carbon emission levels were forecasted based on time series data on carbon emissions from 1949 to 2018 in North America. This study uses 2 forecasting methods, namely SARIMA and LSTM, with the consideration that both methods are considered capable of providing good results. Forecasting results show that the best parameter for ARIMA is (0,1,0) with a MAPE of 4,2482% and SARIMA is [(0,1,0) (1,1,0)12] with a MAPE of 1.995%. Meanwhile, the LSTM method with parameters 1 input, 4 hidden layers, and output 1, it produces a MAPE of 0.540%. This condition makes the LSTM method more optimal for predicting carbon emission levels in the world.

**Key words:** Carbon Emission, Forecasting, LSTM, MAPE, SARIMA.

## I. PENDAHULUAN

Udara merupakan salah satu komponen kehidupan bagi makhluk hidup. Di mana, makhluk hidup menggunakan udara bersih sebagai sumber pernapasan [1]. Berdasar data WHO pada tahun 2018, terdapat lebih dari 90% populasi dunia menghirup udara dengan tingkat polusi yang berbahaya. Bahkan, 3 miliar orang, yang didominasi wanita dan anak-anak masih menghirup asap mematikan. Kondisi ini diakibatkan oleh rendahnya kualitas udara, terutama yang diakibatkan oleh residu hasil pembakaran senyawa karbon, seperti penggunaan kompor dan bahan bakar [2].

Emisi karbon merupakan gas yang dihasilkan dari pembakaran senyawa karbon. Emisi karbon menjadi salah satu penyumbang emisi gas rumah kaca yang menjadi faktor utama terjadinya pemanasan global [3]. Porsi emisi karbon menjadi gas tertinggi dalam emisi gas rumah kaca, yaitu 56%. Berdasarkan laporan *Intergovernmental Panel on Climate Change* (IPCC) tahun 2021, kegiatan manusia menyumbang pemanasan iklim di dunia, terutama yang disebabkan oleh emisi gas rumah kaca dari pembakaran bahan bakar fosil dan penggunaan lahan [4]. Selain itu, meningkatnya populasi manusia di dunia menjadikan produksi emisi rumah kaca khususnya karbon dan karbondioksida meningkat [5]. Berdasarkan data IQAir

tahun 2020, Bangladesh menjadi negara paling berpolusi di dunia dengan nilai rata-rata 77,10  $\mu\text{g}/\text{m}^3$ . Sedangkan untuk negara-negara lain, mayoritas tergolong ke dalam kategori *unhealthy for sensitive groups* [6]. Kondisi tersebut yang menjadikan permasalahan emisi karbon saat ini menjadi salah satu perhatian utama dunia.

Perlu adanya komitmen yang kuat dalam mengatasi tingginya emisi karbon di dunia. Upaya-upaya yang mulai dilakukan oleh negara-negara maju adalah dengan membuat kebijakan untuk mengurangi penggunaan bahan bakar minyak terutama yang bersumber dari fosil sebagai bahan bakar mesin atau kendaraan bermotor [4]. Selain itu, upaya penambahan ruang terbuka hijau juga dilakukan sebagai upaya pengurangan emisi karbon di kota-kota besar [7]. Namun, rencana-rencana tersebut perlu adanya persiapan yang matang untuk memperkirakan sejauh mana implementasi dan upaya penanganan emisi karbon akan dilakukan, sehingga akan mengurangi biaya produksi dan mengurangi terjadinya kegagalan dalam pengimplementasian. Salah satu cara yang dapat dilakukan, yaitu dengan melakukan peramalan keadaan emisi karbon untuk periode selanjutnya.

Terdapat beberapa penelitian yang sebelumnya telah dilakukan oleh Jia Dong pada tahun 2021 untuk melakukan prediksi emisi karbon menggunakan metode dekomposisi dan analisis *decoupling*. Prediksi ini dilakukan untuk menganalisis pembangunan ekonomi di Provinsi Henan, China. Latar belakang lokasi dipilih dengan pertimbangan China menjadi negara dengan konsumsi energi fosil terbesar di dunia [8]. Penelitian lainnya juga dilakukan di China dengan metode *random forest*. Penelitian ini dilakukan oleh Yuan Fang pada tahun 2021 berfokus pada emisi karbon yang dihasilkan dari proses konstruksi bangunan [9]. Sedangkan dalam metode *neural network*, terdapat penelitian yang dilakukan oleh Claudio Maino pada tahun 2021. Penelitian ini menggunakan metode *Deep Neural Network* atau DNN untuk memprediksi kendaraan listrik hibrid dan didapati hasil performa yang sangat baik, yaitu di atas 91% serta rata-rata regresi errornya kurang dari 1% [10].

Dari ketiga penelitian di atas didapatkan permasalahan emisi karbon akan membentuk suatu tren dari tahun ke tahun, terutama pada negara-negara maju dan memiliki

konsumsi kendaraan bermotor yang tinggi. Namun, pada kenyataannya proses prediksi terkadang mengalami hambatan yang diakibatkan dengan proses perhitungan yang rumit dan kurangnya data [9]. Dalam penelitian ini dilakukan peramalan emisi karbon dengan metode *Seasonal Autoregressive Integrated Moving Average* (SARIMA) dan *Long Short Term Memory* (LSTM). Pemilihan kedua metode untuk dibandingkan karena data emisi karbon memiliki tipe data stasioner (bukan acak), sedangkan berdasar beberapa teori metode SARIMA lebih cocok untuk tipe data stasioner (bukan acak). Pemilihan metode pembandingan LSTM dirasa perlu untuk melihat performa dengan menggunakan *Deep Machine Learning*. Data yang digunakan merupakan data sekunder emisi karbon di Amerika Utara dari tahun 1949 hingga 2018. Metode SARIMA dan LSTM memiliki kemampuan yang baik dalam membaca pola trend dalam data, sehingga mampu secara optimum melakukan peramalan pada data emisi karbon [11]

Data terdiri dari 2 (dua) variabel yang akan digunakan dalam proses peramalan, yaitu *year* atau tahun dan  $Y_t$  yang menunjukkan data emisi karbon pada tahun tersebut per Juta Ton. Data dikelompokkan menjadi data *training* untuk melakukan pelatihan dengan model yang dibentuk, serta data *testing* untuk melakukan pengujian model yang buat. Peramalan dilakukan berdasarkan data *testing*. Hasil peramalan mampu dijadikan acuan bagaimana negara-negara mampu mengurangi emisi karbon di masa yang akan datang [12].

## II. TINJAUAN PUSTAKA

### A. Emisi Karbon

Emisi erat kaitannya dengan proses perpindahan suatu zat atau benda. Menurut Cambridge Dictionary, emisi adalah sejumlah gas, panas, cahaya, dan lain-lain yang dikimkan ke udara. Sedangkan emisi karbon adalah gas residu yang dihasilkan dari pembakaran senyawa yang mengandung karbon, seperti CO<sub>2</sub>, solar, LPJ, dan bahan bakar lainnya. Gas residu tersebut akan dipancarkan atau dilepaskan ke udara. Emisi karbon merupakan salah satu jenis sumber pencemaran udara yang mengakibatkan gas rumah kaca dan pemanasan global, di mana emisi karbon dihasilkan dari aktifitas manusia (antropogenik) dari transportasi, sampah, dan penggunaan energi listrik rumah tangga [5].

Dalam mengukur emisi, dapat dilakukan pengukuran jejak karbon. Jejak karbon adalah jumlah emisi gas rumah kaca yang diproduksi oleh suatu entitas, dimana entitas tersebut merupakan organisasi, peristiwa, produk atau individu. Emisi karbon yang dihasilkan berasal dari aktifitas sehari-hari baik aktivitas konsumen maupun produsen [13].

Berdasarkan data IPCC tahun 2021, emisi karbon menduduki porsi tertinggi penyebab terjadinya gas rumah kaca yaitu sebesar 56% [4]. Dalam kondisi yang ekstrim, pencemaran udara dianggap sebagai bahaya terbesar bagi lingkungan yang menyebabkan anomali iklim yang parah

[8]. Bahkan, turut menyumbang sebanyak 7 juta kematian di seluruh dunia per tahunnya [2].

### B. Time Series

*Time series* merupakan sebuah pengumpulan data yang diakumulasikan dalam beberapa waktu tertentu. Metode ini mengasumsikan bahwa yang terjadi di masa sekarang adalah berasal dari apa yang terjadi di masa lalu. Data masa lalu akan diidentifikasi pola historisnya lalu identifikasi apakah terdapat pola yang berulang di dalamnya. Metode *time series* ini merupakan salah satu metode *forecasting* dengan pendekatan kuantitatif [14].

### C. Seasonal Autoregressive Integrated Moving Average (SARIMA)

*Seasonal Autoregressive Integrated Moving Average* (SARIMA) adalah salah satu metode *Autoregressive Integrated Moving Average* (ARIMA) yang ditambahkan parameter seasonal didalamnya. Karena metode SARIMA ini merupakan salah satu pengembangan dari metode ARIMA, maka dasar ilmu dari metode SARIMA sama seperti metode ARIMA. Terdapat metode *Autoregressive* (AR), *Moving Average* (MA), dan juga penggabungan *Autoregressive* dengan *Moving Average*, yaitu *Autoregressive Moving Average* (ARMA). Berikut pada persamaan 1 adalah persamaan dari metode *autoregressive* [15]:

$$Y_t = \theta_0 + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \dots + \theta_p Y_{t-p} - e_t \quad (1)$$

Selain persamaan *autoregressive*, juga terdapat persamaan dari metode *moving average*. Berikut pada persamaan 2 adalah persamaan metode *moving average* [15]:

$$Y_t = \phi_0 + \phi_1 e_{t-1} - \phi_2 Y_{t-2} - \dots - \phi_p Y_{t-p} \quad (2)$$

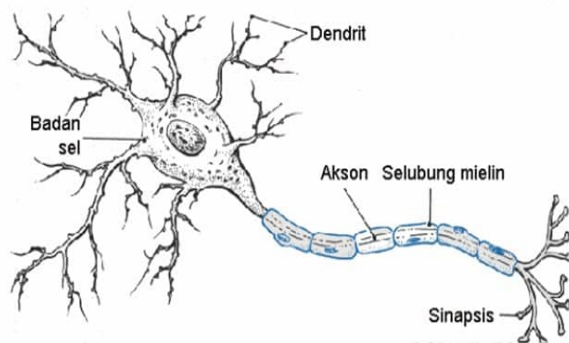
Setelah terdapat persamaan *autoregressive* dan *moving average*, selanjutnya adalah persamaan dari penggabungan antara metode *autoregressive* dan juga *moving average*, yaitu *autoregressive moving average* (ARMA). Berikut pada persamaan 3 adalah persamaan ARMA [15]:

$$Y_t = \gamma_0 + \delta_1 Y_{t-1} - \delta_2 Y_{t-2} - \dots - \lambda_p e_{t-p} \quad (3)$$

### D. Long Short Term Memory (LSTM)

*Long Short Term Memory* atau LSTM merupakan salah satu metode yang memiliki arsitektur yang sama dengan *Recurrent Neural Network* (RNN) [11]. Di mana RNN juga merupakan salah satu metode dalam Jaringan Syaraf Tiruan (JST) yang menerapkan data sekuensial dalam arsitekturnya [16]. Jaringan Syaraf Tiruan atau JST merupakan suatu metode yang diadopsi dari jaringan syaraf biologis untuk mempelajari sesuatu [17]. JST mampu mempelajari suatu permasalahan terutama yang

melibatkan pengelompokan dan pengenalan pola [18]. Proses pembelajaran yang terjadi dalam JST akan membaca pola-pola pada data untuk menggambarkan hubungan yang dinamis antara input dan output. Sistem pembelajaran yang terjadi secara kontinuitas sehingga data yang digunakan akan dieksploitasi secara maksimal untuk mengetahui suatu objek [19]. Proses pembelajaran dalam JST melibatkan dua jenis pembelajaran, yaitu pembelajaran dengan pengawasan (supervised learning) dan pembelajaran tanpa pengawasan (unsupervised learning). Supervised learning memungkinkan output yang berasal dari bobot yang diinginkan. Sedangkan unsupervised learning memungkinkan output dihasilkan dari bobot tanpa melalui acuan tertentu.



Gambar 1. Jaringan Syaraf Biologis (Sumber: dosenpendidikan.co.id)

Dalam arsitektur pengolahan data, JST merupakan multi-layer perceptron yang memiliki 3(tiga) lapisan, yaitu input layer, hidden layer, dan output layer. Berdasar Gambar 1, terdapat 3(tiga) komponen utama dalam jaringan syaraf biologis yang dapat dipelajari oleh JST, yaitu dendrit, akson, badan sel, dan sinapsis. Masing-masing memiliki fungsi yang sama dengan komponen pada arsitektur JST. Hal tersebut diasumsikan bahwa JST mampu mengolah informasi pada komponen-komponen pemrosesan (neuron). TABEL merupakan perbandingan jaringan syaraf biologis dengan JST.

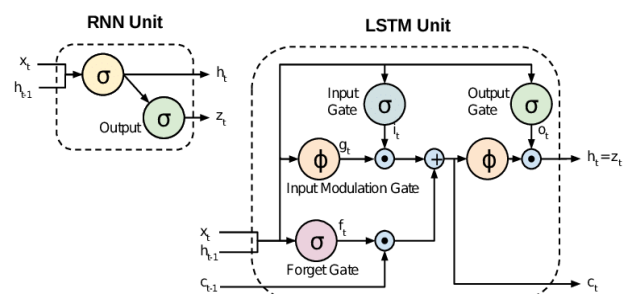
TABEL I. PERBANDINGAN JARINGAN SYARAF BIOLOGIS DENGAN JST

Jaringan Syaraf Biologis	Jaringan Syaraf Tiruan
Dendrit	Input
Akson	Output
Badan sel	Node
Sinapsis	Bobot

Pada umumnya, metode LSTM digunakan untuk menyelesaikan permasalahan-permasalahan yang berkaitan dengan deep learning. Selain itu, LSTM sangat cocok untuk penyelesaian masalah dengan mengklasifikasikan, memproses, dan memprediksi berdasarkan data time-series. Hal tersebut, dikarenakan terdapat kelangkaan durasi data yang tidak diketahui di antara barisan peristiwa penting dalam rangkaian waktu [11].

LSTM akan memproses data yang masuk secara berulang-ulang. Perulangan pada diri sendiri terjadi di

setiap *output* dari *hidden layer*, begitu seterusnya hingga memperoleh hasil *output* yang dirasa paling akurat [16]. LSTM memungkinkan untuk menambahkan *memory cell* yang dapat menyimpan informasi untuk jangka waktu yang lama [20]. Bagian dari *memory cell* dan *gate unit* digunakan sebagai fungsi pengaturan memori pada setiap *neuron* [16]. Arsitektur pada LSTM terdiri atas *memory cell*, *input gate*, *output gate*, dan *forget gate*. LSTM cell akan memasukkan dan menyimpan data dalam jangka waktu tertentu. Secara intuitif, *input gate* yang akan mengontrol sejauh mana informasi baru mampu berjalan ke dalam cell, selanjutnya *forget gate* akan mengontrol sejauh mana informasi tetap di dalam cell, dan *output gate* akan mengontrol informasi yang keluar dalam cell mampu menghitung aktivasi dari model LSTM [11]. Gambar 2 merupakan arsitektur dalam LSTM.



Gambar 2. Arsitektur LSTM (Sumber: stats.stackexchange.com)

Informasi atau data yang masuk dalam *forget gates* akan melalui proses pengolahan sesuai dengan informasi atau data yang ingin disimpan dan *memory cell*. Fungsi aktivasi yang digunakan dapat menggunakan fungsi sigmoid. Persamaan 4 merupakan perhitungan dari nilai pada *input gate* [11].

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{4}$$

Sedangkan untuk persamaan 5 dan 6 digunakan untuk menghitung nilai kandidat dari *input gate* (2 gate) dalam *memory cell* yang menggunakan fungsi sigmoid, yang tujuannya untuk memperbarui informasi dan menyimpan nilai baru dengan menggunakan fungsi aktivasi tahun.

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{5}$$

$$c_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \tag{6}$$

Persamaan 7 merupakan penggabungan nilai yang ada dalam input gate

$$c_t = f_t * c_{t-1} + i_t * c_t \tag{7}$$

*Forget gate* akan menilai *memory cell* dengan *cell gates*. Selanjutnya dilanjutkan ke *output gates*, di mana

dalam *output gates* terdapat 2 *gate* untuk memutuskan nilai yang akan muncul dengan perhitungan persamaan 8 dan 9.

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (8)$$

$$h_t = o_t.tahn(c_t) \quad (9)$$

*E. Pengukuran Performa Model*

Pengukuran performa model dilakukan untuk mengetahui hasil akurasi dari model yang dibuat. Dengan kata lain, pengukuran performa model digunakan untuk mengevaluasi model peramalan. Dalam penelitian ini, pengukuran performa model dilakukan berdasar perhitungan *Mean Percentage Error* (MAPE). MAPE merupakan persentase *error* yang dihasilkan dari nilai rata-rata dari selisih antara data aktual dengan data hasil peramalan. Semakin kecil nilai MAPE dari suatu model, dapat dikatakan model peramalan yang dibuat semakin baik performanya. Tabel II merupakan batasan dari nilai MAPE [21].

TABEL II. BATASAN MAPE

Nilai MAPE	Model Peramalan
<10%	Sangat Baik
10%-20%	Baik
20%-50%	Cukup Baik
>50%	Buruk

Dengan rumus MAPE pada persamaan 10 sebagai berikut:

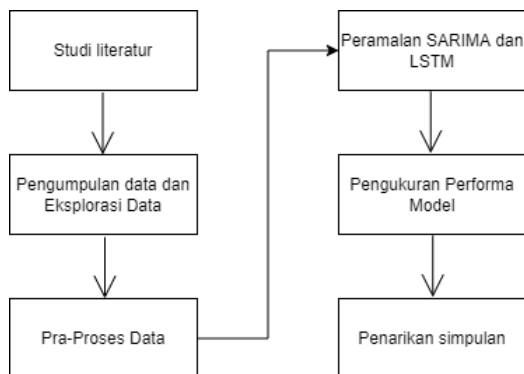
$$MAPE = 100\% * \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (10)$$

di mana:

- MAPE = nilai presentase *error*
- n = jumlah data
- $A_t$  = nilai aktual
- $F_t$  = nilai hasil peramalan

III. METODOLOGI

Pada Gambar 3 dapat dilihat tahapan penelitian yang dilakukan.



Gambar 3. Flow Chart Penelitian

Berikut merupakan penjelasan dari *flow chart* penelitian pada Gambar 3.

*A. Studi Literatur*

Studi literatur adalah kegiatan merumuskan permasalahan yang ada. Selanjutnya setelah menemukan permasalahan, ditetapkan tujuan dari penelitian ini. Studi literatur dilakukan dengan proses telaah beberapa artikel terkait topik dan metode penelitian seperti halnya buku, artikel ilmiah, maupun *website* resmi. Sehingga dapat dijadikan acuan yang terarah terkait topik dan metode yang digunakan.

*B. Eksplorasi Data*

Eksplorasi data adalah kegiatan menyaring data yang akan digunakan untuk penelitian. Selanjutnya pada tahap ini, dipilih variabel yang dibutuhkan dan tidak dibutuhkan. Untuk variabel yang tidak dibutuhkan akan dihapus, dan variabel yang dibutuhkan pada penelitian akan dikumpulkan menjadi dataset yang baru.

*C. Pra-proses Data*

Pra-proses data adalah kegiatan dalam penelitian untuk menentukan pemilihan data *testing* dan data *training*. Pada penelitian ini, *dataset* emisi karbon total terdapat 70 data yang tersimpan dari tahun 1949 hingga tahun 2018. Untuk metode SARIMA, dari total 70 data tersebut, dipilih 70 data untuk data *testing*, dan 11 data *training*. Data *training* diambil dari 11 data terbaru yang ada di dalam dataset emisi karbon tersebut, yaitu dari tahun 2008 hingga tahun 2018. Sedangkan untuk metode LSTM, menggunakan perbandingan dengan 75% merupakan data *training* dan 25% data *testing*.

*D. Peramalan SARIMA*

Untuk peramalan dengan menggunakan metode SARIMA, menggunakan *source code* auto-ARIMA. Pada *source code* tersebut ditetapkan beberapa parameter mulai dari parameter minimum arima, parameter maksimum arima, parameter *differencing*, hingga parameter seasonal. Dengan menggunakan auto-ARIMA dapat secara otomatis mendeteksi parameter model terbaik dari SARIMA tanpa harus mendefinisikannya secara manual dan melakukan *trial and error* untuk model terbaiknya.

*E. Permalan LSTM*

Pada tahap peramalan dengan metode LSTM dilakukan dengan menghitung nilai input *gate* menggunakan persamaan 4. Selanjutnya dilakukan perhitungan kandidat dari *input gate* dalam *memory cell* dengan menggunakan persamaan 5 dan 6 Kemudian dengan persamaan 7 dapat digunakan untuk menggabungkan nilai dalam *input gate*. Dan yang terakhir, untuk memunculkan *output* dari hasil model LSTM dapat dilakukan dengan persamaan 8 dan 9.

F. Pengukuran Performa Model

Setelah dilakukan pemodelan menggunakan metode SARIMA dan LSTM, perlu dilakukan pengukuran performa terhadap model. Pengukuran performa dilakukan untuk mengevaluasi dan mengetahui tingkat akurasi dari model yang dibuat. Hasil peramalan akan dibandingkan dengan data aktual menggunakan perhitungan *Mean Absolute Percentage Error (MAPE)* yang telah dibahas pada sub bab sebelumnya.

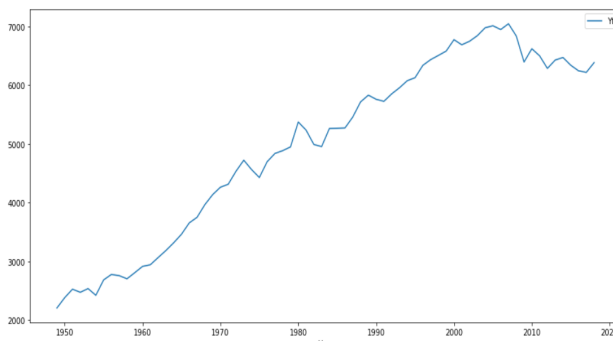
IV. HASIL DAN PEMBAHASAN

A. Raw Data

Data yang digunakan merupakan data sekunder emisi karbon di Amerika Utara pada tahun 1949 hingga 2018. Sebelum proses pemodelan dengan SARIMA dan LSTM, perlu diketahui pola data yang ada di dalam *dataset* emisi karbon. Pada *dataset* emisi karbon memiliki tipe *data time series*, dimana data tersebut dikumpulkan berdasarkan suatu variabel dalam parameter waktu tertentu. Jika dilihat pada Gambar 4, terlihat data emisi karbon terdapat pola tren, yang artinya terjadi peningkatan kadar emisi karbon dari tahun ke tahun. Namun, pada tahun 2006 mulai terdapat penurunan, meskipun tidak signifikan. Setelah terjadi penurunan pada tahun 2006, pola data yang terjadi mulai stabil dan tidak menunjukkan kenaikan yang signifikan. Fenomena data seperti ini perlu dilakukannya differencing untuk menghilangkan tren dan *seasonal* dalam data.

TABEL III. DATA EMISI KARBON AMERIKA UTARA TAHUN 1949-2018

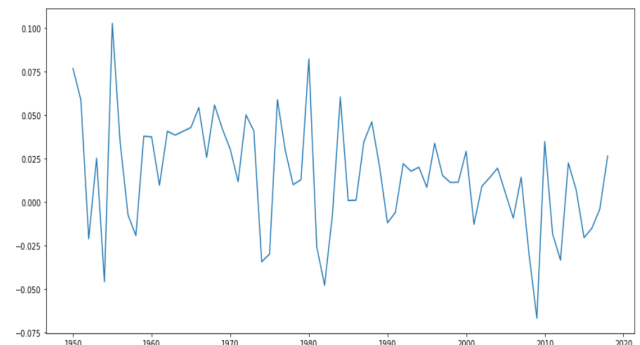
Tahun	Kadar CO2 (per Juta Ton)
1949	2.206
1950	2.382
1951	2.526
...	...
...	...
2016	6.241
2017	6.215
2018	6.381



Gambar 4. Raw Data (Sumber data: kaggle.com)

Setelah mengetahui bahwa pada *raw data* terdapat pola tren, maka yang dilakukan adalah melakukan differencing data. *Differencing* data adalah proses penghitungan selisih

dari nilai observasi. Dengan menghitung nilai selisih tersebut, maka akan dapat membuat data menjadi lebih *stasioner*. Dalam kasus dataset emisi karbon ini, hanya dilakukan *differencing* sebanyak 1 kali, Hal ini disebabkan karena hasil dari *differencing* tersebut telah berhasil membuat raw data menjadi stasioner. Hasil dari *differencing* data terlihat pada Gambar 5. Karena jumlah *differencing* hanya satu kali dan data sudah stasioner, maka parameter *differencing* pada model ARIMA, maksimal adalah 1.



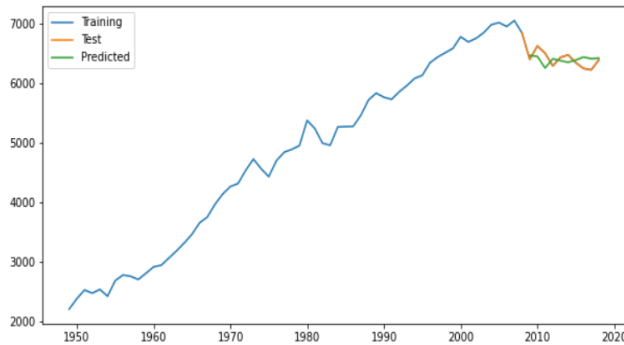
Gambar 5. Differencing Data

B. Peramalan dengan Model SARIMA

Proses penggunaan metode SARIMA diawali dengan melakukan *differencing* 1 kali pada raw data. Setelah melakukan *differencing* 1 kali, mulai terlihat bahwa data tersebut bisa *stasioner*, sehingga parameter *differencing* untuk ARIMA adalah 1. Untuk melakukan pelatihan pada model, dilakukan *training* data dan *testing* data. Pada parameter *training* data dilakukan *setting* parameter sebanyak 70. Hal ini menyesuaikan dengan jumlah data yang terdapat pada dataset emisi karbon. Selain melakukan *training* data, dilakukan juga *test* data. Pada proses *test* data diambil 15% data dari jumlah data keseluruhan dari *dataset*, dan 15% dari total 70 data adalah 11 data. Pada data test tersebut diambil 11 data terbaru, yaitu data pada tahun 2008 hingga 2018.

Setelah melakukan *training* data, yang selanjutnya dilakukan adalah mulai memproses data menggunakan metode auto arima. Lalu didapatkan parameter SARIMA terbaik, yaitu [(0,1,0) (1,1,0)12]. Dengan menggunakan parameter SARIMA [(0,1,0) (1,1,0)12], didapatkan nilai *Akaike Information Criteria (AIC)* sebesar 761.321 dan juga nilai *Mean Absolute Percentage Error (MAPE)* sebesar 1,995%. Setelah itu dilakukan pula pengetesan pada parameter ARIMA tanpa seasonal. Didapatkan parameter ARIMA (0,1,0) dengan *Mean Absolute Percentage Error (MAPE)* sebesar 4,2482%. Pengetesan terhadap metode ARIMA diharapkan mampu memberikan gambaran ketika *seasonal* dalam model dihilangkan, dan diketahui nilai MAPE dengan ARIMA model lebih tinggi. Selain lebih mendapatkan nilai MAPE yang lebih rendah, penggunaan SARIMA lebih dipilih dibanding dengan ARIMA dikarenakan dalam dataraw terdapat pola naik turun (*seasonal*) meskipun tidak signifikan dibandingkan pola trennya.

Berdasarkan hasil MAPE tersebut, dapat disimpulkan bahwa tingkat kesalahan prediksi cukup kecil, sehingga model prediksi tersebut dapat digunakan untuk meramalkan emisi karbon. Hal ini diperkuat dengan perbandingan data *testing*, data *training* dan juga data prediksi yang terdapat pada Gambar 6. Pada Gambar 6 tersebut, terlihat bahwa hasil prediksi tidak terlalu jauh dengan data emisi karbon yang sesungguhnya.

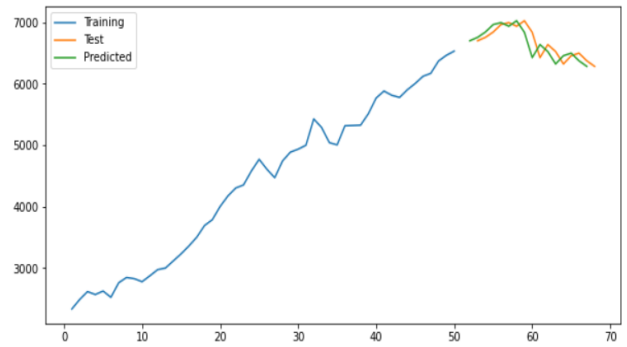


Gambar 6. SARIMA

C. Peramalan dengan Model LSTM

Pada peramalan selanjutnya dilakukan dengan model LSTM perlu dilakukan pembagian antara data *training* dan data *testing*. Pembagian data *training* sebanyak 52 data, dan data *testing* sebanyak 18 data. Dengan begitu proses peramalan dengan model LSTM dapat dilakukan.

Terdapat beberapa parameter yang berkaitan dengan model LSTM ini, seperti *input layer*, *hidden layer*, *output layer*, dan fungsi aktivasi. Parameter yang digunakan dalam pengimplementasian model LSTM ini, yaitu *input* sebanyak 1, *hidden layer* sebanyak 4, dengan *output layer* sebanyak 1. Sedangkan fungsi aktivasi yang digunakan yaitu fungsi aktivasi sigmoid, dimana fungsi aktivasi ini merupakan ciri khas dari model LSTM. Pengujian dilakukan sebanyak 100 kali dengan 1 *batch size*. Gambar 7 merupakan hasil visualisasi dari peramalan menggunakan LSTM, dengan hasil peramalan (dengan garis hijau) mendekati optimum dari data *testing* (dengan garis oranye). Didapatkan nilai MAPE dari LSTM, yaitu 0,540%. Angka tersebut tergolong ke dalam model yang sangat baik dikarenakan nilai MAPE dari LSTM bernilai kurang dari 10% dan mendekati nol.



Gambar 7. LSTM

D. Perbandingan Hasil Peramalan ARIMA, SARIMA dan LSTM

Perbandingan hasil peramalan SARIMA dan LSTM dilakukan berdasar pada nilai akurasi yang dihasilkan dari kedua model. Nilai akurasi didapatkan melalui perhitungan nilai MAPE untuk masing-masing model. Pada TABEL dapat dilihat bahwa nilai MAPE dari model LSTM lebih kecil dibandingkan dengan nilai MAPE dari model SARIMA. Begitu juga dengan

TABEL IV. PERBANDINGAN MAPE

Model	MAPE
ARIMA	4,2482%
SARIMA	1,995%
LSTM	0,540%

Selain berdasarkan nilai MAPE dari kedua model, perbandingan hasil peramalan dapat dilakukan dengan melakukan visualisasi data hasil peramalan. Pada Gambar 8 merupakan visualisasi data dari *raw data* dengan garis baru, hasil peramalan LSTM dengan garis oranye, dan hasil peramalan SARIMA. Pola hasil peramalan dengan LSTM lebih mendekati grafik dari *raw data* dibandingkan dengan grafik hasil peramalan SARIMA. Selain itu juga bentuk grafik dari LSTM lebih mengikuti dari grafik *raw data*.



Gambar 8. Perbandingan ARIMA dan SARIMA

## V. KESIMPULAN

Dari hasil dan pembahasan pada poin sebelumnya, dapat ditarik sebuah kesimpulan yaitu metode LSTM lebih optimal jika digunakan untuk meramalkan kadar emisi karbon dibandingkan dengan metode SARIMA. Hal ini dibuktikan dengan besarnya *Mean Absolute Percentage Error* atau MAPE yang di dapatkan pada saat melakukan proses *testing* data dengan menggunakan data *testing*. Dapat diketahui bahwa MAPE dari metode LSTM adalah 0,540% sedangkan MAPE dari metode SARIMA adalah 1,995%. Apabila nilai MAPE semakin mendekati 0, maka model yang dibuat semakin optimum. Maka dari itu berdasarkan MAPE yang didapat, LSTM adalah metode yang lebih optimal dibandingkan dengan metode SARIMA.

Selain itu, bukti bahwa model dari metode LSTM lebih optimum dibandingkan dengan metode SARIMA dapat dilihat melalui grafik pada Gambar 8. Hal ini disebabkan karena grafik peramalan yang didapatkan dari metode LSTM cenderung mengikuti pola dari *raw data*. Sedangkan metode SARIMA memiliki pola tersendiri, tidak mengikuti pola dari *raw data* yang ada. Sehingga hal ini menyebabkan hasil prediksi dari SARIMA memiliki jarak yang lebih besar dengan *raw data* dibandingkan metode LSTM yang memiliki jarak lebih kecil dengan *raw data*.

Saran untuk penelitian mendatang adalah untuk menyesuaikan metode yang digunakan dengan dataset yang ada. Karena dalam penelitian ini memiliki banyak data sehingga mendukung metode LSTM. Pada metode LSTM semakin banyak data historis yang digunakan, maka akan semakin akurat pula hasil dari peramalan tersebut. Namun sebaliknya, pada metode SARIMA, semakin banyak *dataset* yang digunakan, cenderung tidak akurat. Karena pada metode SARIMA, yang terambil untuk data prediksi hanya beberapa data terakhir saja. Maka dari itu, sebelum memilih metode lebih baik disesuaikan dulu dengan dataset yang akan digunakan.

## DAFTAR PUSTAKA

- [1] J. Abidin and F. A. Hasibuan, "Pengaruh Dampak Pencemaran Udara Terhadap Kesehatan Untuk Menambah Pemahaman Masyarakat Awam Tentang Bahaya Dari Polusi Udara," *Prosiding Seminar Nasional Fisika Universitas Riau*, pp. 1-7, 2019.
- [2] WHO, "9 out of 10 people worldwide breathe polluted air, but more countries are taking action," 31 12 2020. [Online]. Available: <https://www.who.int/news/item/02-05-2018-9-out-of-10-people-worldwide-breathe-polluted-air-but-more-countries-are-taking-action>.
- [3] D. Labiba and W. Pradoto, "Sebaran Emisi CO2 dan Implikasinya terhadap Penataan Ruang Area Industri di Kendal," *Jurnal Pengembangan Kota*, pp. 164-173, 2018.
- [4] X. Han, "Spatiotemporal characteristics of carbon emissions in energy-enriched areas and the evolution of regional types," *Energy Report*, 2021.
- [5] I. W. J. Gobel, "Sebaran Spasial Emisi Gas Karbon Dioksida (CO2) pda Kawasan Pemukiman di Kecamatan Singkil Kota Manado," *Jurnal Spasial*, 2019.
- [6] IQAir, "Negara Paling Berpolusi di Dunia," IAAir, 31 12 2020. [Online]. Available: <https://www.iqair.com/id/world-most-polluted-countries>. [Accessed 16 12 2021].
- [7] C. H. Edyanto, "Emisi KArbon Sebagai Dasar Implementasi Penyediaan RUang Terbuka Hijau di DKI Jakarta," *Peneliti Pusat Teknologi - Sumberdaya Lahan Wilayah dan Mitigasi Bencana - Deputi Bidang Pengembangan Kekayaan Alam BPPT*, 2013.
- [8] J. Dong, "Decomposition of carbon emission and its decoupling analysis and prediction with economic development: A case study of industrial sectors in Henan Province," *Journal of Cleaner Production*, 2021.
- [9] Y. Fang, X. Lu and H. Li, "A random forest-based model for the prediction of construction-stage carbon emissions at the early design stage," *Jurnal of Clear Production*, 2021.
- [10] C. Maino, D. Misul, A. D. Mauro and E. Spessa, "A deep neural network based model for the prediction of hybrid electric vehicles carbon dioxide emissions," *Energi and AI*, 2021.
- [11] A. Khumaidi, "Pengujian Algoritma Long Short Term Memory untuk Prediksi Kualitas Udara dan Suhu Kota Bandung," 2020.
- [12] P. Hasanah and I. Fitria, "Prediksi Emisi CO2 dengan Analisis Runtun Waktu," *SPECTA Journal of Technology*, pp. 21-25, 2017.
- [13] N. Aqualdo, Eriyati and T. Indrawati, "Penyeimbangan Lingkungan Akibat Pencemaran Karbon yang Ditimbulkan Industri Warung Internet di Kota Pekanbaru," *Jurnal Ekonomi*, vol. 20, pp. 1-11, 2012.
- [14] L. Yuliana, "Analisis Perencanaan Penjualan Dengan Metode Time Series (Studi Kasus Pada PD. Sumber Jaya Aluminium)," *Jurnal Mitra Manajemen*, pp. 780-789, 2019.
- [15] B. Hendrawan, "Penerapan Model ARIMA Dalam Memprediksi IHSG," 2013.
- [16] A. Farhah, A. L. Prasasti and M. W. Paryasto, "Implementasi Recurrent Neural Network dalam Memprediksi Kepadatan Restoran Berbasis LSTM," *Jurnal Media Informatika Budidarma*, pp. 524-531, 2021.
- [17] H. D. Bhakti, "Aplikasi Artificial Neural Network (ANN) untuk Memprediksi Masa Studi Mahasiswa Program Studi Teknik Informatika Universitas Muhammadiyah Gresik," *Jurnal Eksplor Informatika*, pp. 88-95, 2019.
- [18] I. F. Ningsih and R. Salambue, "Klasifikasi Kematangan Buah Sawit dengan Jaringan Syaraf Tiruan Metode Perceptron," *Repositori Universitas Riau*, 2019.
- [19] Syukri and Samsuddin, "Pengujian Algoritma Artificial Neural Network (ANN) untuk Prediksi Kecepatan

- Angin," *Jurnal Nasional Komputasu dan Teknologi Informasi*, pp. 43-47, 2018.
- [20] A. S. Girsang, "Long Short Term Memory (LSTM)," Binus University, 1 1 2018. [Online]. Available: <https://mti.binus.ac.id/2019/12/02/long-short-term-memory- lstm/>. [Accessed 16 12 2021].
- [21] W. Anggreani, "Analisis Kinerja Metode ANFIS untuk Peramalan Kasus Demam Berdarah di Kabupaten Malang," *Jurnal SIsfo*, 2018.