

# Perbandingan Metode *Long Short Term Memory* (LSTM) DAN *Genetic Algorithm-Long Short Term Memory* (GA-LSTM) Pada Peramalan Polutan Udara

Oleh: Cita Meliana<sup>1</sup> Rochdi Wasono<sup>2</sup> M Al Haris<sup>3</sup>  
<sup>1,2,3</sup>Univeristas Muhammadiyah Semarang  
email: [melianacita@gmail.com](mailto:melianacita@gmail.com)

Article history	Abstract
Submission :	Air is one of the primary needs of living things. To maintain air quality, it can be done by forecasting levels of air pollutants using historical data. Prediction of air pollutant levels can be done using the Long Short Term Memory (LSTM) and Genetic Algorithm-Long Short Term Memory (GA-LSTM) methods. LSTM is built for sequential data processing such as time series data which is intended to avoid long-term dependency problems. Genetic Algorithm (GA) is an optimization technique based on the principles of genetics and natural selection which in this study is used to find the optimal window size and number of units. Based on the research results, the best model was obtained for forecasting air pollutant levels using LSTM with the number of neurons 50 and epoch 30 with a MAPE value of 4.54%. Meanwhile, data forecasting of air pollutant levels using the GA-LSTM method obtained the best model with window size 33 and number of units 9 with a MAPE value of 1.83%. From these results, it can be concluded that the GA-LSTM method is better at modeling air pollutant levels in the city of Jakarta.
Revised :	
Accepted :	
<b>Keyword:</b> <i>Long Short Term Memory, Genetic Algoritihm, Polutan Udara, Jakarta</i>	

## PENDAHULUAN

Udara merupakan salah satu komponen yang paling penting untuk kelangsungan makhluk hidup terutama manusia. Menurut Hesam (2005) dalam Naddafi dkk (2006), kebutuhan udara bagi manusia lebih utama daripada kebutuhan terhadap makanan dan air. Rata-rata kebutuhan udara orang dewasa adalah 15 kg/hari, sedangkan kebutuhan makanan dan air masing-masing sebesar 1.5 kg/hari dan 2.5 kg/hari. Manusia dapat terus hidup tanpa makanan selama lima minggu dan tanpa air selama lima hari, tetapi tidak lebih dari beberapa menit tanpa udara. Namun, kondisi udara di perkotaan saat ini sudah berubah dari batas normal. Hal ini disebabkan oleh polusi udara seperti dari sisa hasil pembakaran mesin kendaraan bermotor atau emisi gas buang, kegiatan industri atau kegiatan lain yang mengakibatkan udara tercemar.

Pencemaran udara dikenal sebagai masalah lingkungan yang terasosiasi dengan

wilayah perkotaan di seluruh dunia. Menurut Laporan *State of Global Air* tahun 2018 yang diterbitkan oleh *Health Effects Institute* di Amerika dan *Institute for Health Metrics and Evaluation's Global Burden of Disease Project*, polusi udara merupakan penyebab terbesar ke-6 yang dapat memperpendek usia manusia atau kematian dini. Sektor transportasi sebagai tulang punggung aktivitas masyarakat berkontribusi besar pada pencemaran udara karena mayoritas kendaraan bermotor masih menggunakan bahan bakar fosil. Menurut Badan Pusat Statistik, jumlah kendaraan bermotor di Indonesia yang mencakup mobil penumpang, mobil bis, mobil barang dan sepeda motor dalam kurun waktu 10 tahun terakhir (2007-2017) mengalami pertumbuhan yang sangat tinggi yakni sebesar 153%. Sepeda motor mendominasi dengan 82% di tahun 2017. Seiring meningkatnya jumlah kendaraan bermotor yang beredar di masyarakat menyebabkan meningkatnya polutan udara yang berupa emisi gas buang. Emisi gas buang

merupakan gas yang dibuang oleh kendaraan bermotor dari sisa hasil proses pembakaran bahan bakar di dalam mesin. Emisi dari kendaraan bermotor yang terus-menerus terpapar di udara menghasilkan berbagai macam polutan, salah satunya adalah *particulate matter* 10 (PM10) sebagai akibat proses pembakaran mesin yang tidak sempurna serta partikel lepas (Syahrani, 2006).

Dinas Lingkungan Hidup DKI Jakarta mencatat konsentrasi PM10 di udara Jakarta selama lima tahun terakhir mencapai 60,8 mikrogram per meter kubik. Angka tersebut jauh di atas jumlah standar konsentrasi udara menurut Badan Kesehatan Dunia (WHO) yakni 20 mikrogram per meter kubik dalam jangka waktu satu tahun (Tresnasari, 2018). Berdasarkan hal tersebut perlu dilakukan sebuah peramalan mengenai bagaimana kadar polutan udara ke depannya. Peramalan (*forecasting*) adalah suatu teknik analisis perhitungan untuk memperkirakan kejadian di masa yang akan datang dengan menggunakan pengalaman di masa lampau (Sunayang, 2013). Peramalan bisa menjadi dasar bagi para pembuat kebijakan untuk membuat perencanaan jangka pendek, menengah maupun jangka panjang bagi organisasi komersial maupun tidak. Salah satu algoritma yang dipakai dalam peramalan yang memanfaatkan *Artificial Intelligence* dan berbasis data runtun waktu adalah *Long Short Term Memory* (LSTM) yang merupakan jenis dari *Recurrent Neural Network* (RNN).

Penelitian terdahulu menggunakan LSTM diantaranya Zheng dkk (2017), dalam penelitian ini model LSTM dibandingkan dengan beberapa model peramalan lain yaitu dengan SARIMA, NARX, SVR, dan NNETAR. Dihasilkan bahwa model LSTM mampu mengungguli metode peramalan lain dengan menghasilkan nilai RMSE dan MAPE yang lebih rendah. Nilai RMSE yang dihasilkan yaitu sebesar 0,0702, sementara nilai MAPE yang dihasilkan sebesar 0,0535%. Aulia (2020) tentang prediksi harga Ethereum dengan menggunakan metode LSTM menunjukkan bahwa model prediksi terbaik diperoleh dengan jumlah neuron 50 dan *max epoch* 500 serta nilai MAPE sebesar 1.69 %. Terdapat juga penelitian oleh Sugiarto (2017) mengenai optimasi *Artificial Neural Network* menggunakan *Genetic Algorithm* pada prediksi *approval credit card* dengan menggunakan algoritma *neural network* mendapatkan hasil

peningkatan dari 85.42% menjadi 87.82% hasil optimasi dengan *Genetic Algorithm*. Chung dkk (2018) melakukan penelitian tentang prediksi mengenai harga saham KOSPI dengan menggunakan *Genetic Algorithm-Long Short Term Memory* (GA-LSTM) yang menghasilkan bahwa *time window size* yang diperoleh dengan optimasi GA sebesar 10 dan menggunakan dua *hidden layer* dengan jumlah *node* 15 dan 7.

Dalam penelitian ini akan dilakukan *hybrid* antara GA dengan LSTM dalam memprediksi kualitas udara kota Jakarta yang selanjutnya akan dibandingkan antara metode LSTM dengan GA-LSTM. Melalui penelitian ini, diharapkan dapat diperoleh model terbaik yang dapat memberikan hasil prediksi yang akurat mengenai kualitas udara di Kota Jakarta dan dapat dijadikan sebagai salah satu upaya untuk meningkatkan sistem pemantau kualitas udara yang ada di Kota Jakarta. Dari penelitian ini diharapkan para pemangku kepentingan khususnya di bidang pengendalian polusi lingkungan, dapat menggunakan metode ini sebagai acuan pengambilan keputusan dalam upaya mengendalikan polusi udara untuk tahun-tahun yang akan datang.

## LANDASAN TEORI

### *Particulate Matter* 10 $\mu\text{m}$ (PM10)

Partikel PM10 terdiri dari partikel kompleks berukuran 0,1  $\mu\text{m}$ –10  $\mu\text{m}$ , mencakup semua ukuran virus (0,1  $\mu\text{m}$ –1  $\mu\text{m}$ ) dan bakteri (0,5  $\mu\text{m}$ –5 $\mu\text{m}$ ) (Lai *et al.*, 2009). PM10 merupakan salah satu oksidan pencemar yang dapat dihisap oleh saluran pernapasan (Yusnabeti *et al.*, 2010). Konsentrasi debu di atmosfer didominasi oleh PM10. Konsentrasi PM10 yang tinggi dapat menyebabkan gangguan kesehatan terutama pada manusia seperti efek akut (batuk, sesak napas, nyeri dada, iritasi pada mata, detak jantung tidak beraturan, menurunkan fungsi paru, memperparah penyakit paru dan jantung) dan efek kronis (gangguan pada sistem saraf dan pembuluh darah dan memicu kematian dini pada orang dengan penyakit paru dan jantung) (Wulandari, 2016:680).

### *Artificial Neural Network* (ANN)

*Artificial Neural Networks* (ANN) atau Jaringan Saraf Tiruan (JST) merupakan suatu model kecerdasan yang diilhami dari struktur otak manusia yang kemudian diimplementasikan menggunakan program komputer yang mampu menyelesaikan

sejumlah proses perhitungan selama proses pembelajaran berlangsung. ANN yaitu suatu model penalaran yang didasarkan pada otak manusia. ANN terdiri dari sejumlah prosesor sangat sederhana dan saling berhubungan yang disebut neuron (Kurniawansyah, 2018). Neuron memiliki karakteristik yang sama dalam ANN, terdiri dalam kelompok-kelompok yang di sebut layer. Neuron-neuron yang ada dalam satu layer terhubung dalam layer-layer lainnya yang berdekatan. Kekuatan hubungan antara neuron yang berdekatan dipresentasikan dalam kekuatan hubungan atau bobot. (Dharma, Putera, & Ardana, 2011).

### **Recurrent Neural Network (RNN)**

RNN adalah sejenis jaringan saraf tiruan yang memiliki kemampuan untuk melihat korelasi tersembunyi yang terjadi pada data yang aplikasinya digunakan pengenalan suara, pemrosesan bahasa alami, dan prediksi deret waktu. RNN sangat baik untuk masalah pemodelan urutan dengan beroperasi pada informasi input serta jejak informasi yang diperoleh sebelumnya karena koneksi berulang (Tian & dkk, 2018).

*Recurrent Neural Network (RNN)* terdiri dari unit input, unit output, dan unit tersembunyi. Model RNN pada dasarnya memiliki aliran informasi satu arah dari unit input ke unit tersembunyi, dan sintesis aliran informasi satu arah dari unit tersembunyi sementara sebelum ke unit tersembunyi waktu saat ini. Unit tersembunyi dapat dilihat sebagai penyimpanan seluruh jaringan, yang mengingat informasi dari ujung ke ujung (Yin, 2017). Secara teori, RNN mampu menangani ketergantungan jangka panjang. Namun, dalam aplikasi praktis, RNN tidak dapat menyimpan informasi sebelumnya dengan baik ketika interval waktu lama karena masalah gradien menghilang. Untuk mengatasi kelemahan ini dan meningkatkan kinerja RNN, sebuah tipe khusus arsitektur RNN yang disebut LSTM diusulkan oleh Hochreiter & Schmidhuber (1997) (Tian & dkk, 2018).

### **Long Short Term Memory (LSTM)**

*Long Short Term Memory (LSTM)* menggunakan salah satu bentuk RNN yang paling umum yang dimaksudkan untuk menghindari masalah ketergantungan jangka panjang dan sesuai untuk memproses serta memprediksi deret waktu. Model LSTM

menyaring informasi melalui struktur gerbang untuk mempertahankan dan memperbarui keadaan sel memori. Struktur pintunya mencakup gerbang input, *forget gate*, dan gerbang output. Setiap sel memori memiliki tiga lapisan sigmoid dan satu lapisan tanh.

*Forget gate* dalam unit LSTM menentukan informasi status sel mana yang dibuang dari model. Fungsi utama dari *forget gate* adalah untuk merekam seberapa banyak status sel  $C_{t-1}$  dari waktu sebelumnya dicadangkan ke status sel  $C_t$  dari waktu saat ini.

*Input gate* atau gerbang input menentukan berapa banyak input jaringan waktu saat ini  $x_t$  dicadangkan ke dalam status sel, yang mencegah konten tidak signifikan dari memasuki sel memori. Gerbang input memiliki dua fungsi, pertama adalah menemukan keadaan sel yang harus diperbarui, nilai yang akan diperbarui dipilih oleh lapisan sigmoid dan untuk memperbarui informasi ke  $b$  diperbarui ke keadaan sel.

Gerbang output mengontrol seberapa banyak keadaan sel saat ini dibuang. Informasi keluaran pertama-tama ditentukan oleh lapisan sigmoid, dan kemudian keadaan sel diproses oleh tanh dan dikalikan dengan keluaran lapisan sigmoid untuk mendapatkan bagian keluaran akhir.

### **Genetic Algorithm (GA)**

Algoritma Genetika merupakan suatu teknik optimasi yang didasarkan pada prinsip genetik dan seleksi alam. Dalam Algoritma Genetika populasi terbentuk dari banyak individu yang berkembang sesuai aturan seleksi spesifik dengan memaksimalkan *fitness* (Haupt dan Haupt, 2004). Algoritma ini juga digunakan untuk mendapatkan nilai global optimum dengan cara melakukan perulangan atau iterasi pada konsep evolusi darwin.

### **Root Mean Square Error (RMSE)**

Untuk menentukan model terbaik dilakukan dengan RMSE (*Root Mean Square Error*) untuk setiap model dengan nilai RMSE terkecil menyatakan model terbaik. RMSE dirumuskan sebagai berikut:

$$RMSE = \sqrt{\frac{1}{m \times n} \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} [f(i, j) - g(i, j)]^2} \quad (1)$$

### **Mean Absolute Percentage Error (MAPE)**

*Mean Absolute Percentage Error* (MAPE) adalah nilai absolute dari persentase error data terhadap mean. Persamaannya ditulis seperti di bawah ini:

$$MAPE = \frac{\sum_{i=1}^n \frac{|\hat{Y}_i - Y_i|}{Y_i}}{n} * 100\% \quad (2)$$

Tabel 1. Kriteria MAPE

MAPE	Pengertian
<10%	Kemampuan Peramalan Sangat Baik
10%-20%	Kemampuan Peramalan Baik
20%-50%	Kemampuan Peramalan Cukup
>50%	Kemampuan Peramalan Buruk

## METODE PENELITIAN

### Sumber Data

Data yang digunakan pada penelitian ini merupakan data sekunder yang diperoleh dari stasiun pemantauan udara ambien di Dinas Lingkungan Hidup DKI Jakarta. Dinas Lingkungan Hidup DKI Jakarta memiliki 5 SUF (Stasiun Pemantauan Udara Ambien), namun dalam penelitian ini hanya fokus pada 1 stasiun yaitu di Bundara HI (SUF1). Data pada penelitian ini adalah data rata-rata parameter polutan udara per jam yang berupa particulate matter 10 (PM10) pada tanggal 1 Januari 2020 sampai dengan 30 November 2020 dengan jumlah data sebanyak 7705 data. Data yang diperoleh akan dibagi menjadi data training dan data testing. Menurut Agustina (2010) data training dan data testing dibagi dengan komposisi 80% data *training* dan 20% data *testing*.

### Variabel dan Struktur Data

Tabel 2. Variabel Data

Tanggal dan Waktu	Wisman	Keterangan
01/01/2020 01:00	Y <sub>1</sub>	Training
01/01/2020 02:00	Y <sub>2</sub>	-
-	-	-
25/09/2020 23:00	Y <sub>6164</sub>	Training
26/09/2020 01:00	Y <sub>6165</sub>	Testing
-	-	-
30/11/2020 23:00	Y <sub>7705</sub>	Testing

### Langkah Penelitian

Langkah-langkah dalam penelitian ini yaitu :

1. *Preprocessing data* dengan melihat ada atau tidaknya *missing value*. Jika terdapat *missing value* dilakukan imputasi data atau mengganti data dengan nilai rata-rata periode sebelum dan sesudahnya (Evriyanto, 2004).
2. Identifikasi data untuk mengetahui karakteristik data kadar polutan udara di Kota Jakarta diantaranya :
  - a. Melakukan analisis deskriptif pada data kadar polutan udara di Kota Jakarta.
  - b. Membuat grafik *time series* dari data kadar polutan udara kota Jakarta.
3. Melakukan normalisasi pada data penelitian dengan mengubah data aktual menjadi nilai dengan range interval [0,1] menggunakan *min-max scaling*.
4. Membagi data menjadi data *training* dan data *testing* dengan proporsi untuk data training sebesar 80% dan data testing sebesar 20%.
5. Melakukan analisis LSTM dengan langkah-langkah sebagai berikut :
  - a. Membuat model *supervised learning problem*.
  - b. Menentukan jumlah *neuron* pada *hidden layer*
  - c. Menentukan jumlah *epoch*
  - d. Prediksi data
  - e. Akurasi data dengan menggunakan MAPE
6. Melakukan optimasi LSTM menggunakan GA dengan langkah-langkah sebagai berikut:
  - a. Inisiasi parameter GA dengan menentukan jumlah populasi, panjang kromosom, jumlah generasi, *crossover rate* dan *mutation rate*.
  - b. Melakukan *training* jaringan LSTM dengan menentukan jumlah *neuron* pada *hidden layer* yang diperoleh dari GA dan menentukan jumlah *epoch*.
  - c. Evaluasi nilai *fitness* untuk setiap kromosom yang terbentuk. Lalu, dilihat apakah kriteria terminasi apakah terpenuhi atau tidak. Jika kriteria terminasi terpenuhi dapat dilanjutkan ke langkah 5d. Jika kriteria terminasi tidak terpenuhi harus dilakukan *genetic search* dan kembali ke langkah 5a.
  - d. Optimasi parameter GA pada model jaringan LSTM dengan melakukan testing data untuk meramalkan kadar polutan udara dengan jaringan LSTM yang terbentuk dengan parameter GA.

e. Akurasi data dengan menggunakan MAPE.

7. Melakukan perbandingan ketepatan peramalan pada metode LSTM dan GA-LSTM.

## HASIL PENELITIAN dan PEMBAHASAN

### Statistika Deskriptif

Berdasarkan data kadar polutan udara selama 2020 tidak terdapat trend kenaikan atau penurunan kadar polutan udara yang signifikan sepanjang tahun 2020. Pada tanggal 6 November 2020 merupakan rata-rata kadar polutan udara tertinggi yaitu sebesar 152,7  $\mu\text{g}/\text{m}^3$  dengan kadar tertinggi pada pukul 05:00 sebesar 322,76  $\mu\text{g}/\text{m}^3$ . Sedangkan rata-rata kadar polutan udara terendah terjadi pada tanggal 1 Maret 2020 yaitu sebesar 18,9  $\mu\text{g}/\text{m}^3$ .

### Preprocessing Data

Pada tahap ini, data mentah dalam bentuk excel digabungkan menjadi satu file dataset dan disimpan dalam format excel (.xlsx). Jika terdapat *missing value* atau nilai yang bukan angka (---) dan nilai 0 diganti dengan nilai rata-rata kadar polutan udara sebelum dan sesudahnya.

### Pebentukan Data Training dan Data Testing

Data penelitian dibagi menjadi data *training* dan data *testing* dengan perbandingan 80% untuk data *training* atau sebanyak 6164 dan 20% untuk data *testing* atau sebanyak 1541.

### Analisis Long Short Term Memory (LSTM)

Sebelum masuk dalam proses *machine learning*, dataset harus diubah ke dalam model *supervised learning problem*. *Supervised learning problem* adalah meringkai ulang masalah pembelajaran diawasi dari data masukan dan keluaran. Caranya adalah menggunakan langkah waktu sebelumnya sebagai variabel input dan menggunakan langkah waktu berikutnya sebagai variabel output. Tabel 4.3 merupakan kadar polutan udara pada tanggal 01/01/2020 dari pukul 02.00 sampai dengan 08.00. Jika diubah ke dalam model *supervised learning problem* maka bisa dilihat pada tabel 3. Dalam tabel 3 dapat diartikan bahwa  $\text{var}(t)$  adalah nilai yang akan diprediksi yang masih terpengaruh nilai  $\text{var}(t-1)$ .

Tabel 3. Hasil *Supervised Learning Problem*

Tanggal/Jam	$\text{Var}(t-1)$	$\text{Var}(t)$
01/01/2020 02:00	0,929391	0,558587
01/01/2020 03:00	0,558587	0,558587
01/01/2020 04:00	0,558587	0,558587
01/01/2020 05:00	0,558587	0,187783
01/01/2020 06:00	0,187783	0,137402
01/01/2020 07:00	0,137402	0,180772
01/01/2020 08:00	0,180772	0,177142

### Penentuan Jumlah Neuron dan Epoch

Dalam menentukan jumlah *neuron*, tidak ada aturan yang mengatur banyaknya *neuron* yang digunakan, sehingga nilai *neuron* dapat di pakai peneliti dengan terlebih dahulu melakukan percobaan sampai mendapatkan hasil yang optimal dalam prediksi yang bisa di lihat dari nilai error. Begitu juga dengan *epoch*, *epoch* adalah langkah yang dilakukan pada proses pembelajaran *neural network*, dimana besarnya *epoch* yang telah ditetapkan akan mempengaruhi besaran proses pembelajaran dan berhenti tepat pada nilai *epoch* yang telah ditentukan tersebut.

Pada penelitian ini jaringan dibentuk dengan jumlah *neuron* pada *hidden layer* yang akan digunakan untuk percobaan yaitu 10, 20, 30, 40 dan 50. Sedangkan untuk *epoch* akan digunakan adalah 10, 30, 50, 80, 100, 150 dan 300. Untuk mengetahui jumlah *neuron* dan *epoch* yang tepat dapat dilihat nilai *loss* yang terkecil, dimana nilai *loss* merupakan nilai RMSE serta menggunakan optimasi Adam. Tidak ada aturan dalam menentukan jumlah *neuron* dan *epoch* sehingga jumlah *neuron* dan *epoch* didapatkan melalui percobaan sampai mendapatkan yang paling optimal dalam memprediksi *time series*.

Tabel 4. Hasil Training LSTM

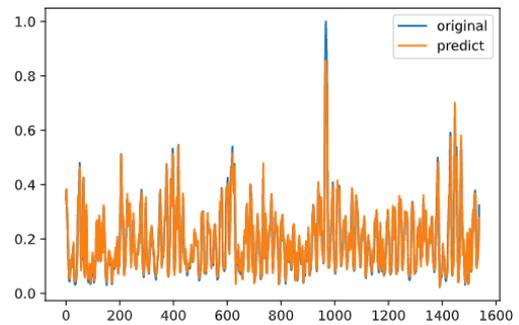
Jumlah Neuron	Epoch	RSME
10	10	12.66468
	30	12.66468
	50	12.477546
	80	12.570128
	100	12.61102
	150	12.648782
20	300	12.687255
	10	12.513792
	30	12.436901
	50	12.43043
	80	12.513219

	100	12.539952
	150	12.559182
	300	12.577465
30	10	12.510565
	30	12.66468
	50	12.42547
	80	12.509002
	100	12.536481
	150	12.556524
	300	12.574208
40	10	12.482178
	30	12.407989
	50	12.443811
	80	12.518735
	100	12.533073
	150	12.541876
	300	12.552052
50	10	12.4886
	30	<b>12.401656</b>
	50	12.478155
	80	12.563479
	100	12.564858
	150	12.547664
	300	12.545849

Tabel 4 menunjukkan model LSTM terbaik dengan menggunakan *neuron 50* dan *epoch 30* diperoleh nilai error terendah yaitu 12.401656. Jumlah epoch merepresentasikan lamanya proses pembelajaran yang dilakukan terhadap jaringan yang sedang diobservasi. Jumlah *epochs* yang terlalu sedikit mengakibatkan jaringan yang terbentuk bersifat terlalu general, berarti kemampuan jaringan dalam mengenali pola terlalu sedikit atau bahkan tidak ada sama sekali. Sedangkan jumlah *epochs* yang terlalu banyak akan mengakibatkan jaringan mengalami kondisi *overfit* (jaringan bersifat terlalu spesifik terhadap data pelatihan), itu tampak pada tabel 4 sehingga hasil terbaik tidak berada pada nilai *epoch* terbesar, melainkan *neuron 50* dan *epoch 30*, arsitektur inilah yang akan dipakai dalam prediksi kadar polutan udara.

#### Peramalan Kadar Polutan Udara dengan LSTM

Setelah diperoleh arsitektur terbaik untuk melakukan peramalan dengan menggunakan data testing 20 % serta dengan *neuron 50* dan *epoch 30* maka di dapatkan grafik perbandingan data aktual kadar polutan udara dengan data prediksi kadar polutan udara seperti pada gambar 1.



Gambar 1. Plot Perbandingan Data Prediksi dengan Data Aktual

Peramalan kadar polutan udara di Kota Jakarta diperoleh akurasi 94,46% berdasarkan nilai *Mean Absolute Percentage Error (MAPE)*. Dengan nilai MAPE 4,54% menunjukkan bahwa hasil prediksi sangat baik.

#### Analisis Genetic Algorithm-Long Short Term Memory (GA-LSTM)

Pendekatan hybrid antara jaringan Long Short Term Memory (LSTM) dengan Genetic Algorithm (GA) dimaksudkan untuk menemukan jumlah *window size* dan *number of unit* yang sesuai untuk LSTM. LSTM menggunakan informasi masa lalu selama proses pembelajarannya sehingga *window size* dan *number of unit* yang dipilih dengan tepat akan menghasilkan peramalan yang lebih akurat.

#### Inisialisasi Parameter GA

Tahapan pertama dalam proses ini adalah menentukan populasi awal. Diperlukan populasi awal sebelum optimasi dilakukan. Populasi ini merupakan kumpulan dari kromosom-kromosom yang berisikan solusi untuk jumlah *window size* dan *number of unit*. Pembentukan kromosom dilakukan secara random dengan bilangan biner. Dalam penelitian ini ditentukan:

Populasi	= 4
Panjang kromosom	= 10
Maksimum generasi	= 4
Crossover rate	= 0.7
Mutation rate	= 0.15

#### Training Jaringan LSTM dan Evaluasi Fitness

Tahapan kedua yaitu merancang arsitektur jaringan LSTM yang akan digunakan. Dalam penelitian ini *epoch* yang digunakan

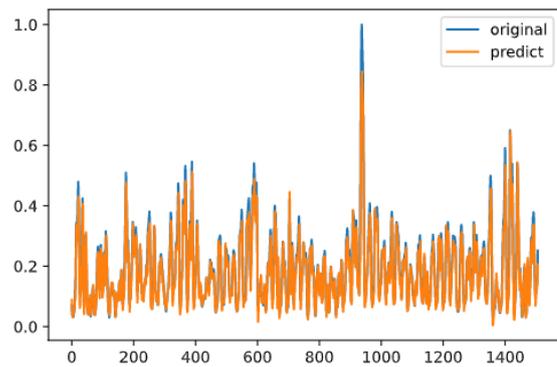
sebesar 5 dan ukuran *window size* dan *number of units* ditentukan dengan GA. Berbagai ukuran *window size* dan *number of units* akan diterapkan dalam jaringan LSTM untuk mengevaluasi nilai *fitness* atau kesesuaian GA. RMSE digunakan untuk menghitung nilai *fitness* setiap kromosom dan nilai RMSE terkecil dianggap sebagai solusi yang optimal. Apabila kriteria terminasi dalam satu generasi sudah terpenuhi maka proses GA dihentikan dan dilanjutkan ke model peramalan, jika belum terpenuhi maka seluruh proses GA akan diulangi lagi. Berikut hasil nilai *fitness* setiap kromosom GA yang digunakan pada jaringan LSTM:

Tabel 5. Hasil *Training* GA-LSTM

<i>Window Size</i>	<i>Number of Units</i>	RMSE
39	5	0.043527
54	15	0.031746
3	1	0.066858
37	10	0.038635
37	9	0.042111
39	9	0.035946
39	5	0.039134
37	9	0.038320
37	10	0.037673
37	9	0.042507
37	10	0.039856
20	14	0.032143
38	10	0.039044
37	9	0.038268
33	9	<b>0.036905</b>
23	12	0.039212

### Peramalan Kadar Polutan Udara dengan GA-LSTM

Setelah diperoleh arsitektur LSTM terbaik dengan GA untuk melakukan peramalan, dengan menggunakan data *testing* 20 % serta dengan 33 *window size*, 9 *number of unit* dan *epoch* 5 maka di dapatkan grafik perbandingan data aktual kadar polutan udara dengan data prediksi kadar polutan udara seperti pada gambar dibawah ini.



Gambar 2. Plot Perbandingan Data Prediksi dengan Data Aktual

Gambar 2 menunjukkan perbandingan antara data aktual dengan data prediksi kadar polutan udara. Peramalan kadar polutan udara dengan menggunakan GA-LSTM ini menghasilkan akurasi sebesar 98,27% dengan nilai MAPE 1,83% yang menunjukkan hasil prediksi sangat baik.

### Perbandingan LSTM dan GA-LSTM

Dari hasil yang didapatkan, maka perbandingan ketepatan peramalan dari metode LSTM dan GA-LSTM sebagai berikut :

Tabel 6. Perbandingan Ketepatan Peramalan

Metode	MAPE
LSTM	4,54
GA-LSTM	1,83

Berdasarkan tabel 6, peramalan kadar polutan udara dengan menggunakan GA-LSTM menghasilkan akurasi yang lebih baik dibandingkan LSTM dengan nilai MAPE sebesar 1,83%.

### SIMPULAN dan SARAN

#### Simpulan

Berdasarkan penelitian diatas dapat disimpulkan bahwa metode LSTM dapat diterapkan pada data kadar polutan udara dengan menggunakan *neuron* sebesar 50 dan *epoch* sebesar 30 dan diperoleh akurasi ketepatan prediksi sebesar 96,46% dengan nilai MAPE 4,54%. Untuk metode GA-LSTM diperoleh *window size* dan *number of unit* yang optimal untuk LSTM sebesar 33 dan 9 serta didapatkan nilai akurasi ketepatan peramalan sebesar 98,27% dengan nilai MAPE 1,83%. Dari hasil perbandingan ketepatan peramalan

antara LSTM dan GA-LSTM dapat diketahui bahwa GA-LSTM lebih baik untuk peramalan kadar polutan udara dengan tingkat akurasi peramalan sebesar 98,27%.

### Saran

Berdasarkan hasil penelitian yang sudah dilakukan, saran yang dapat diberikan untuk penelitian selanjutnya adalah optimasi GA pada LSTM dapat digunakan untuk mencari arsitektur atau parameter LSTM yang lain selain ukuran *window size* dan *number of units*. Kedua, dalam penelitian selanjutnya dapat menggunakan optimasi yang lain dalam mengoptimasi arsitektur LSTM atau yang lainnya kemudian dibandingkan.

### Daftar Pustaka

- Aulia, Nilda. (2020). Prediksi Harga Ethereum Berdasarkan Informasi Blockchain Menggunakan Metode Long Short Term Memory. *Skripsi*. Universitas Islam Indonesia. Yogyakarta.
- Badan Pusat Statistik. Jumlah Kendaraan Bermotor. Tersedia: <https://www.bps.go.id/linkTableDinamis/view/id/1133>
- Chung, Hyejung & Shin, Kyung-shik. (2018). Genetic Algorithm-Optimized Long Short-Term Memory Network for Stock Market Prediction. *Sustainability*. 10. 3765. 10.3390/su10103765.
- Health Effects Institute. (2018). State of Global Air 2018. Special Report. Boston, MA:Health Effects Institute.
- Naddafi, K., Nabizadeh, R., Soltanianzadeh, R., Ehrampoosh, M.H. (2006). *Evaluation of Dustfall in The Air of Yazd*. Iran. J Environ. Health. Sci. Eng. 3(3):161-168.
- Olah, C. (2015, Agustus 27). Understanding LSTM Networks. Retrieved from <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- Shin, K.S.; Lee, Y.J. (2002). A genetic algorithm application in bankruptcy prediction modeling. *Expert Syst. Appl.* 23, 321–328.
- Syahrani A. 2006. “Analisis Kinerja Mesin Bensin Berdasarkan Hasil Uji Emisi,” SMARTek, Volume 4, Nomor 4.
- Tian, C., & dkk. (2018). A Deep Neural Network Model for Short-Term Load Forecast Based on Long Short-Term Memory Network and Convolutional Neural Network. *Energies*, 3-4. <https://www.mdpi.com/1996-1073/11/12/3493/pdf>.
- Tresnasari S., Budiyo dan Suhartono. (2018). Gambaran Pola Pencemar Udara di Wilayah Sekitar Bundaran Hotel Indonesia Tahun 2017. *Jurnal Kesehatan Masyarakat Volume 6, Nomor 6, Edisi Oktober UNDIP*.
- Usnabeti, Wulandari, R., Luciana, R., (2010). PM10 dan Infeksi Saluran Pernapasan Akut pada Pekerja Industri Mebel. *MAKARA KESEHATAN* 14, 25–30.
- Wei, W. W. S. (2013). *Oxford Handbooks Online Time Series Analysis* (Vol. 2). <https://doi.org/10.1093/oxfordhb/9780199934898.013.0022>
- World Health Organization. (2018). World health statistics 2018: monitoring health for the SDGs. *Sustainable development goals*, Geneva.
- Wu EMY, Kuo SL. (2013). *A Study on the Use of a Statistical Analysis Model to Monitor Air Pollution Status in an Air Quality Total Quantity Control District*. *Atmosphere*. 4:349-364.doi:10.3390/atmos4040349.
- Zheng, J., & dkk. (2017). Electric Load Forecasting in Smart Grid Using LongShort-Term-Memory based. 51st Annual Conference on Information Sciences and Systems (CISS), doi: 10.1109/CISS.2017.7926112, 1-6.

