

Prediksi Kualitas Udara dengan Menggunakan Metode Long Short-Term Memory dan Artificial Neural Network

I Gusti Ayu Nandia Lestari¹, I Nyoman Dwi Aysna Mahendra²

Institut Teknologi dan Bisnis STIKOM Bali

e-mail: ¹nandialestari@stikom-bali.ac.id, ²dwiarysna97@gmail.com

Diajukan: 29 Desember 2022; Direvisi: 3 Februari 2023; Diterima: 8 Februari 2023

Abstrak

Kualitas indeks udara di Jakarta masuk kategori tidak sehat karena memiliki kualitas udara pada konsentrasi PM_{2.5} mencapai 36,2 $\mu\text{g}/\text{m}^3$, nilai ini tertinggi di Indonesia. Oleh karena itu, dilakukan penelitian untuk memprediksi kualitas udara khususnya PM_{2.5} yang dapat menyebabkan gangguan kesehatan pada manusia atau makhluk hidup lainnya. Pada penelitian ini menerapkan dua metode yaitu Long Short-Term Memory dan Artificial Neural Network. Dataset yang digunakan dalam penelitian didapatkan dari Portal Satu Data yang dipublikasikan oleh Dinas Lingkungan Hidup DKI Jakarta. Dataset ini memiliki beberapa atribut, namun dalam penelitian ini hanya menggunakan atribut PM_{2.5}, PM₁₀, CO, SO₂, NO₂ dan O₃. Nilai koefisien korelasi antara atribut target dengan atribut lainnya adalah PM_{2.5} dengan PM₁₀ yaitu 0.91, PM_{2.5} dengan SO₂ yaitu 0.01, PM_{2.5} dengan CO yaitu 0.29, PM_{2.5} dengan O₃ yaitu 0.50, dan PM_{2.5} dengan NO₂ yaitu 0.47. Berdasarkan hasil pengujian dengan metode LSTM menunjukkan hasil yang lebih baik dari ANN dengan nilai RMSE 23.311, MAE 19.391 dan MAPE 0.176. Sedangkan pemodelan menggunakan metode ANN mendapatkan hasil RMSE 24.673, MAE 21.338, MAPE 0.194.

Kata kunci: LSTM, ANN, PM_{2.5}, ISPU, Prediksi.

Abstract

The air quality index in Jakarta falls into the unhealthy category due to air quality with PM_{2.5} concentration reaching 36.2 $\mu\text{g}/\text{m}^3$, the highest value in Indonesia. Therefore, research has been conducted to predict air quality, especially PM_{2.5}, which can cause health problems for humans or other living creatures. This study applies two methods Long Short-Term Memory and Artificial Neural Network. The dataset used in the research was obtained from the Satu Data Portal published by the Dinas Lingkungan Hidup Jakarta Government. This dataset has several attributes, but this study only uses PM_{2.5}, PM₁₀, CO, SO₂, NO₂, and O₃ attributes. The correlation coefficient values between the target attribute and other attributes are as follows, PM_{2.5} with PM₁₀ is 0.91, PM_{2.5} with SO₂ is 0.01, PM_{2.5} with CO is 0.29, PM_{2.5} with O₃ is 0.50, and PM_{2.5} with NO₂ is 0.47. Based on testing results with the LSTM method, it shows better results than ANN with RMSE values of 23.311, MAE of 19.391, and MAPE of 0.176. Meanwhile, modeling using the ANN method yielded RMSE of 24.673, MAE of 21.338, and MAPE of 0.194.

Keywords: LSTM, ANN, PM_{2.5}, ISPU, Prediction.

1. Pendahuluan

Pencemaran udara di perkotaan tidak dapat dihindari karena berbagai aktivitas masyarakat seperti transportasi, industri, dan konstruksi infrastruktur. Seperti halnya ibu kota Indonesia yaitu DKI Jakarta, yang telah menjadi kota besar dengan jumlah penduduk melebihi 10 juta jiwa, sehingga permasalahan lingkungan termasuk pencemaran udara menjadi sangat penting untuk diperhatikan. Sering kali penduduk daerah tersebut merasakan dampak pencemaran udara terhadap kesehatan yang sangat serius. Menurut WHO pencemaran udara telah berkontribusi sebesar 7.6% kematian di seluruh dunia, dan tidak kurang 7 juta kematian dini setiap tahun, akibat penyakit terkait pencemaran udara baik di dalam ruang (*indoor air pollution*) maupun di luar ruangan. Bahkan 99% penduduk dunia pada tahun 2022 diperkirakan telah terpapar pencemaran udara yang menyebabkan risiko berbagai penyakit, termasuk stroke, jantung, gangguan paru-paru dan kanker [1].

Kualitas indeks udara (AQI) di Jakarta masuk kategori tidak sehat karena memiliki kualitas udara pada konsentrasi PM_{2.5} mencapai 36,2 $\mu\text{g}/\text{m}^3$ di mana angka tersebut menunjukkan bahwa dalam satu meter kubik udara pada lokasi tersebut terdapat 36,2 mikrogram partikulat udara dengan ukuran kurang

dari 2,5 mikrometer. Hal ini tentu sangat berbahaya karena DKI Jakarta merupakan pusat aktivitas dan menjadi kota terpadat di Indonesia. Nilai Air Quality Index (AQI) dihitung berdasarkan dari enam jenis polutan utama diantaranya PM_{2.5} partikulat udara dengan ukuran lebih kecil dari 2.5 mikrometer, PM₁₀ partikulat udara dengan ukuran lebih kecil dari 10 mikrometer, Karbon Monoksida (CO), Asam Belerang (SO₂), Nitrogen Dioksida (NO₂), Ozon (O₃) permukaan tanah. Keenam zat polutan ini menjadi tolak ukur dalam menentukan taraf udara pada Indeks Standar Pencemaran Udara (ISPU) [2].

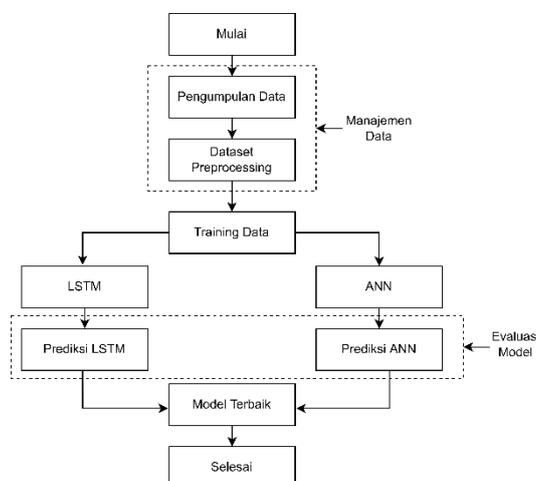
Prediksi data terhadap kualitas udara dapat memperkuat upaya dalam mencegah polusi udara. Pengembangan model prediksi dengan parameter data Indeks Standar Pencemaran Udara (ISPU) di DKI Jakarta, dapat menghasilkan keputusan – keputusan yang baik berdasarkan analisis data yang akurat dalam pencegahan terhadap polusi udara [3]. Pendekatan Machine Learning sering dipakai untuk memprediksi kualitas udara, terutama dalam memprediksi tingkat konsentrasi gas polutan seperti PM_{2.5}. Penelitian tersebut mencoba beberapa metode seperti Long Short Term Memory dan Linier Regression untuk memprediksi konsentrasi PM_{2.5}. Hasil menunjukkan bahwa Long Short Term Memory memiliki kinerja lebih baik dalam memprediksi konsentrasi gas polutan PM_{2.5} [4]. Metode Long Short Term Memory juga digunakan dalam membangun model prediksi kualitas udara dari data Chennai pada tahun 2022. Prediksi PM_{2.5} dengan menggunakan metode Long Short Term Memory pada tahap pelatihan menunjukkan nilai R², untuk tahap pengujian sebesar 0,632 dan 0.570, dengan hal tersebut menunjukkan bahwa metode Long Short Term Memory cocok digunakan dalam memprediksi PM_{2.5} [5]. Pendekatan Machine Learning lainnya yaitu Artificial Neural Network dimana metode ini dapat memecahkan masalah dengan melakukan training data yang besar. Artificial Neural Network memiliki kemampuan untuk mentoleransi kesalahan sehingga dapat menghasilkan prediksi yang baik.

Berdasarkan penjabaran tersebut, penelitian ini bertujuan membangun model prediksi kualitas udara dengan menggunakan metode Long Short-Term Memory (LSTM) dan Artificial Neural Network (ANN). Data yang digunakan untuk pembuatan model didapatkan dari halaman web Satu Data yang diterbitkan oleh Dinas Lingkungan Hidup DKI Jakarta. Data ini diukur dari Stasiun Pemantauan Kualitas Udara (SPKU), sehingga data yang diperoleh sudah memiliki standar khusus dalam proses pengukurannya. Oleh karena itu, dalam penelitian ini tidak dijelaskan secara rinci mengenai proses pengukuran data dan jenis perangkat sensor yang digunakan ataupun kalibrasi dari perangkat sensor tersebut. Kemudian, dengan penerapan metode LSTM dan ANN, diharapkan dapat memberikan kontribusi yang signifikan dalam memprediksi kualitas udara dengan akurat. Prediksi yang tepat sangat penting dalam konteks pencegahan pencemaran udara, sehingga pemerintah dapat merencanakan kebijakan yang lebih efektif dan efisien untuk mengurangi dampak buruk dari polusi udara terhadap kesehatan masyarakat dan lingkungan. Selain itu, hasil dari penelitian ini diharapkan dapat memberikan wawasan yang lebih mendalam tentang dinamika polusi udara, khususnya di DKI Jakarta.

2. Metode Penelitian

2.1. Gambaran Umum Penelitian

Pada bagian ini, dijelaskan gambaran umum dari penelitian yang dilakukan, yang mencakup proses untuk pembuatan model prediksi kualitas udara. Tahapan-tahapan tersebut merangkum langkah-langkah yang dilakukan untuk menghasilkan model terbaik menggunakan berbagai metode.



Gambar 1. Gambaran Umum Penelitian.

Pada Gambar 1 menunjukkan gambaran umum penelitian yang dimulai dari tahap pengumpulan data. Dalam tahap ini pengumpulan data dilakukan dengan mencari data pada halaman web mengenai open data polusi udara di Jakarta ibukota Indonesia. Data yang didapatkan diproses agar dapat dijadikan data untuk pembuatan model prediksi polusi udara. Sebelum ke tahap melakukan proses pembuatan model dataset ini harus di preprosesing dulu untuk memastikan data yang digunakan untuk pembuatan model sudah layak. Proses preproseing dengan menghilangkan beberapa parameter yang tidak diperlukan dan menghapus data yang kosong.

Setelah proses preprocessing dilanjutkan ke tahap untuk training data dengan menggunakan dua metode yaitu LSTM dan ANN. Masing-masing metode ini akan menghasilkan prediksi yang selanjutnya dievaluasi untuk mengetahui nilai *error* dari model tersebut. Hasil model dari metode LSTM dan ANN akan dibandingkan untuk mengetahui model yang terbaik.

2.2. Manajemen Data

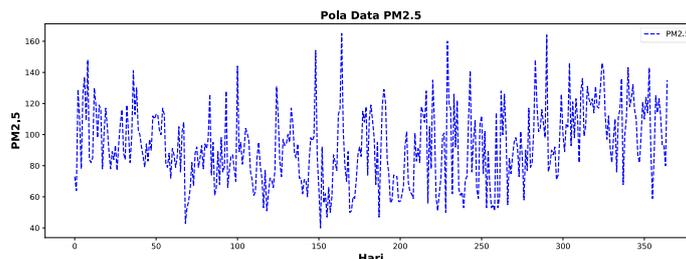
Data yang digunakan dalam penelitian ini disediakan oleh Dinas Lingkungan Hidup DKI Jakarta yang dapat diunduh pada halaman web Satu Data DKI Jakarta. Data Indeks Standar Pencemaran Udara (ISPU) diukur melalui Stasiun Pemantauan Kualitas Udara (SPKU) di provinsi DKI Jakarta [6]. Dalam data ini menggambarkan kondisi mutu udara ambien di lokasi tertentu dengan angka tanpa satuan yang menunjukkan dampak terhadap kesehatan manusia, nilai estetika dan makhluk hidup tertentu. Pada penelitian ini data ISPU yang disediakan dalam dataset diperoleh informasi bahwa pengumpulan data dimulai dari 1 Januari 2022 hingga 31 Desember 2022. Dataset ini memiliki 365 baris data yang menunjukkan banyaknya data dengan 13 kolom atau atribut. Ini menunjukkan jika Dinas Lingkungan Hidup DKI Jakarta melakukan pengukuran data setiap hari atau setiap 24 Jam.

Susunan atribut data yang membentuk dataset yang diunduh pada halaman Satu Data ini seperti *periode, bulan, tanggal, stasiun, pm_10, pm25, sulfur_dioksida, karbon_monoksida, ozon, nitrogen_dioksida, max, parameter_pencemar_kritis dan kategori*. Secara lengkap penjelasan mengenai masing-masing atribut dataset diperlihatkan pada Tabel 1.

Tabel 1. Atribut Penyusun Dataset.

Atribut	Deskripsi
Periode	Periode dalam sebulan
Bulan	Bulan pengambilan data
Tanggal	Tanggal pengambilan data
Stasiun	Lokasi pengambilan data
PM_10	Nilai pengukuran untuk parameter PM 10 Mikron
PM_2.5	Nilai pengukuran untuk parameter PM 2.5 Mikron
Sulfur_dioksida	Nilai pengukuran untuk parameter sulfur dioksida (SO ₂)
Karbon_monoksida	Nilai pengukuran untuk parameter karbon monoksida (CO)
Ozon	Nilai pengukuran untuk parameter ozon (O ₃)
Nitrogen_dioksida	Nilai pengukuran untuk parameter nitrogen dioksida (NO ₂)
Max	Nilai tertinggi hasil pengukuran dari beberapa parameter
Parameter_pencemar_kritis	Nama parameter dengan nilai tertinggi
Kategori	Kategori hasil pengukuran

Berdasarkan atribut dataset yang ditampilkan dalam Tabel 1, peneliti menggunakan hanya enam atribut untuk pembuatan model, adapun atribut tersebut yaitu, PM_10, PM_2.5, SO₂, CO, O₃ dan NO₂. Kemudian, peneliti membagi jumlah data sebanyak 80% (292 data) sebagai data *training* dan 20% (73 data) sebagai data uji atau *testing*. Target dalam penelitian ini adalah memprediksi kualitas polusi udara PM_{2.5} dan pola data untuk PM_{2.5} selama satu tahun diperlihatkan pada Gambar 2.



Gambar 2. Pola Data PM2.5 Selama 1 Tahun.

Langkah selanjutnya dataset ini diproses untuk mengetahui nilai koefisien korelasi. Untuk menghitung nilai koefisien (r) menggunakan rumus 1 [7]. Perhitungan nilai koefisien ini akan menunjukkan korelasi data antara atribut yang ada pada dataset. Pentingnya mengetahui korelasi ini karena tidak semua atribut dalam dataset relevan digunakan untuk memprediksi target (PM_2.5), ini artinya dengan memilih atribut yang sesuai akan maka dalam melatih model menjadi lebih efektif.

Hasil perhitungan nilai koefisien dataset diperlihatkan pada Tabel 2. Dimana, nilai (r) yang memiliki hasil positif ($0 < r < 1$) memiliki korelasi dengan target sedangkan hasil negatif ($-1 < r < 0$) tidak memiliki korelasi dengan target. Berdasarkan hasil yang diperlihatkan pada Tabel 2, terlihat jika target (PM_2.5) memiliki korelasi yang positif terhadap semua atribut. Adapun nilai korelasi antara PM_2.5 dengan PM_10 yaitu 0.91, PM_2.5 dengan SO2 yaitu 0.01, PM_2.5 dengan CO yaitu 0.29, PM_2.5 dengan O3 yaitu 0.50, dan PM_2.5 dengan NO2 yaitu 0.47.

$$r = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \sqrt{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}} \tag{1}$$

Dimana, r adalah nilai koefisien korelasi, $x_i y_i$ merupakan nilai masing dari variabel x dan y , sedangkan n merupakan jumlah sample dari data tersebut.

Tabel 2. Nilai Koefisien Korelasi Dataset.

	PM_10	SO2	CO	O3	NO2	PM_2.5
PM_10	1	0.1	0.3	0.47	0.53	0.91
SO2	0.1	1	0.22	-0.24	0.16	0.01
CO	0.3	0.22	1	-0.08	0.55	0.29
O3	0.47	-0.24	-0.08	1	0.06	0.50
NO2	0.53	0.16	0.55	0.06	1	0.47
PM_2.5	0.91	0.01	0.29	0.50	0.47	1

2.3. Long Short-Term Memory (LSTM)

LSTM merupakan bagian metode dari deep learning yang merupakan jenis dari Recurrent Neural Network (RNN). Metode ini dapat digunakan untuk melakukan prediksi data untuk masa depan. LSTM memiliki keunggulan untuk dapat memproses data dalam jangka waktu yang panjang atau data dalam bentuk time series. Pada Gambar 3 menunjukkan arsitektur dari LSTM, di mana secara umum memiliki tiga gate utama yaitu input gate, forget gate dan output gate. Kemudian persamaan untuk metode LSTM berdasarkan arsitekturnya dijelaskan pada rumus 2 – 9 [8].

$$F_t = \text{sigmoid}([H_{t-1}, X_t] \cdot W_f + b_f) \tag{2}$$

$$I_t = \text{sigmoid}([H_{t-1}, X_t] \cdot W_i + b_i) \tag{3}$$

$$\tilde{C}_t = \tanh([H_{t-1}, X_t] \cdot W_c + b_c) \tag{4}$$

$$C_t = (F_t \cdot C_{t-1}) + (I_t \cdot \tilde{C}_t) \tag{5}$$

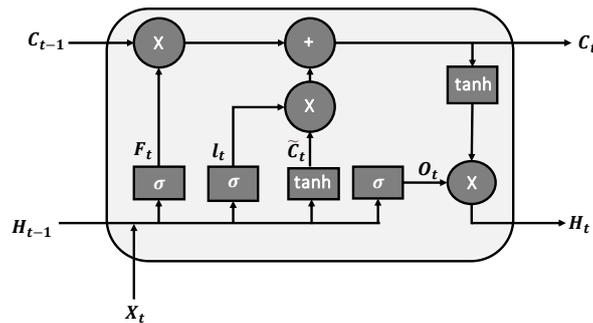
$$O_t = \text{sigmoid}(W_o \cdot [H_{t-1}, X_t] + b_o) \tag{6}$$

$$H_t = O_t \cdot \tanh(C_t) \tag{7}$$

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \tag{8}$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{9}$$

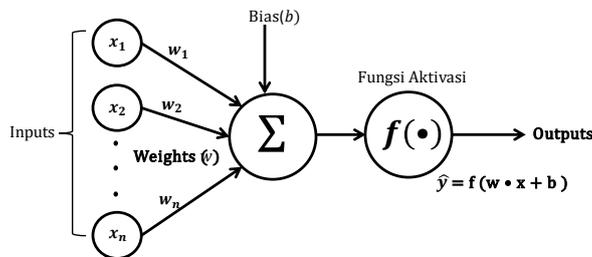
Dimana, F_t merupakan *forget gate*, H_{t-1} adalah nilai dari *hidden state* sebelumnya, X_t merupakan data yang diinputkan, W adalah bobotnya dan b adalah nilai biasnya. I_t ini merupakan *input date*, \tilde{C}_t adalah nilai kandidat dan C_t merupakan *cell state*. Terakhir, O_t merupakan nilai dari *gate output*.



Gambar 3. Arsitektur LSTM.

2.4. Artificial Neural Network (ANN)

ANN atau jaringan saraf tiruan merupakan metode untuk mengelola informasi dengan cara kerja meniru seperti saraf biologis manusia [9]. Arsitektur dari ANN terdiri dari neuron yang saling terhubung satu dengan yang lainnya dengan sistem yang kompleks dan juga data non-linear. Layer dari ANN terdiri dari tiga bagian yaitu layer input, hidden layer dan output seperti yang ditunjukkan pada Gambar 4 [10][11]. Hidden layer menerima data dari layer input dan hidden layer ini bisa banyak tergantung dari data yang diproses. Hasil dari masing-masing hidden layer ini disebut aktivasi.



Gambar 4. Arsitektur Artificial Neural Networks.

2.5. Evaluasi Model

Proses evaluasi model menggunakan berbagai metrik pengukuran untuk masing-masing model. Evaluasi ini dilakukan bertujuan untuk mengetahui sejauh mana performa model yang dihasilkan pada setiap metode. Adapun perhitungan evaluasi yang dilakukan sebagai berikut, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Squared Error (MSE) dan Mean Absolute Percentage Error (MAPE). Persamaan untuk masing-masing teknik evaluasi dijelaskan pada rumus 10 – 13 [12][13].

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \tag{10}$$

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \tag{11}$$

$$MSE = \frac{\sum_{i=1}^n |y(i) - \tilde{y}_i|^2}{n} \tag{12}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \tilde{y}_i|}{y_i} \times 100\% \tag{13}$$

Di mana, n adalah jumlah dari data *sample*, y_i merupakan nilai aktual dari data dan \tilde{y}_i merupakan nilai prediksi.

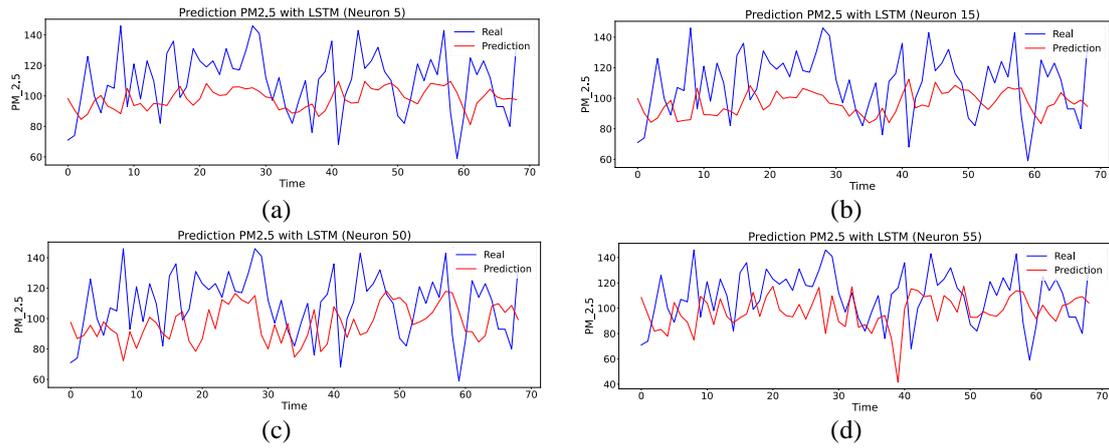
3. Hasil dan Pembahasan

Pada bagian ini menjelaskan hasil dari penelitian untuk pembuatan model prediksi kualitas udara dengan metode LSTM dan ANN. Dalam pemodelan dengan metode LSTM dan ANN salah satu parameter yang paling banyak diatur adalah jumlah neuronnya. Pada penelitian ini dilakukan pengujian dengan mengatur jumlah neuronnya kemudian diamati hasil akhir yang diperoleh untuk masing-masing model. Selain itu, untuk meningkatkan hasil pemodelan dengan data dalam bentuk *time series* juga harus memperhatikan banyaknya data yang digunakan untuk memprediksi data kedepannya atau yang disebut dengan *lookback*. Dalam penelitian ini nilai *lookback* ditentukan sebanyak 4. Nilai *lookback* 4 ini menunjukkan jika empat buah data saat ini digunakan untuk memprediksi satu data kedepan. Pada dataset ini setiap data memiliki interval 1 hari maka 1 *lookback* artinya 1 hari maka, jika 4 *lookback* menunjukkan 4 hari data saat ini digunakan untuk memprediksi nilai PM2.5 satu hari kedepannya.

Berdasarkan hasil pengujian menunjukkan jika metode LSTM lebih sesuai digunakan dibandingkan dengan ANN untuk prediksi kualitas udara di DKI Jakarta. Hal ini dapat dilihat seperti pada Table 3 dengan metode LSTM dan Tabel 4 dengan metode ANN. Pada pengujian tersebut menggunakan metode LSTM dengan jumlah neuron 5 mendapatkan hasil RMSE 23.311, MAE 19.391 dan MAPE 0.176. Sedangkan pemodelan menggunakan metode ANN dengan jumlah neuron 5 mendapatkan hasil RMSE 24.673, MAE 21.338, MAPE 0.194. Selain itu, hasil gambaran plotting antara nilai aktual dengan prediksi diperlihatkan pada Gambar 5 untuk metode LSTM, dan Gambar 6 untuk metode 6. Plotting gambar ini menampilkan sebanyak empat model terbaik dari beberapa kombinasi jumlah neuron yang digunakan pada metode LSTM dan ANN.

Tabel 3. Tingkat Kinerja Model Prediksi PM 2.5 dengan LSTM.

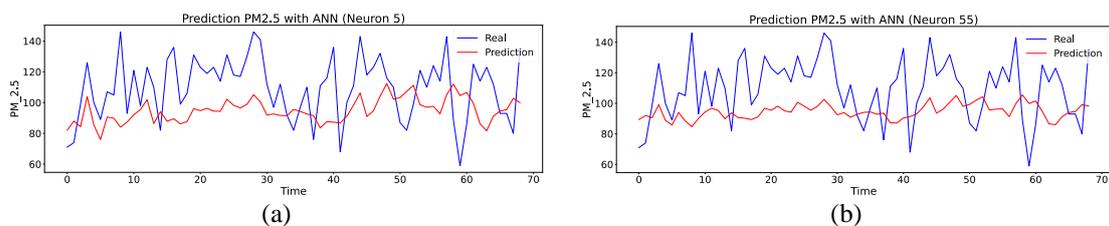
Target	Neuron	Lookback	RMSE (µm)	MSE (µm)	MAE (µm)	MAPE (µm)
PM 2.5	5	4	23.311	543.420	19.391	0.176
PM 2.5	10	4	25.182	634.151	21.169	0.192
PM 2.5	15	4	23.966	574.347	20.123	0.183
PM 2.5	20	4	25.571	653.890	21.289	0.190
PM 2.5	25	4	26.443	699.223	22.139	0.200
PM 2.5	30	4	26.319	692.684	22.356	0.202
PM 2.5	35	4	25.871	669.299	21.365	0.193
PM 2.5	40	4	102.600	10526.676	100.617	0.916
PM 2.5	45	4	26.247	688.913	22.121	0.201
PM 2.5	50	4	25.018	625.894	20.641	0.192
PM 2.5	55	4	25.169	633.487	21.274	0.191
PM 2.5	60	4	31.081	966.009	26.185	0.235
PM 2.5	65	4	30.861	952.399	25.871	0.235
PM 2.5	70	4	26.931	725.253	22.348	0.204
PM 2.5	75	4	28.381	805.484	23.744	0.213
PM 2.5	80	4	27.839	775.004	22.864	0.217
PM 2.5	85	4	27.389	750.143	23.260	0.214
PM 2.5	90	4	25.301	640.116	20.998	0.200
PM 2.5	95	4	27.152	737.224	22.184	0.206
PM 2.5	100	4	29.916	894.983	24.913	0.235

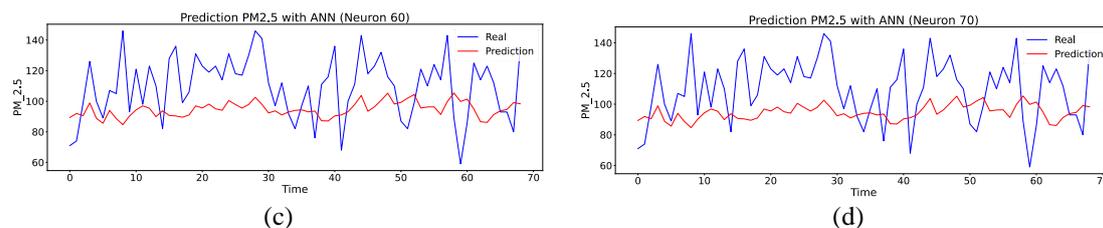


Gambar 5. Perbandingan Data Aktual dan Prediksi dengan LSTM
 (a) Neuron 5 RMSE 23.311 , (b) Neuron 15 RMSE 23.966, (c) Neuron 50 RMSE 25.018,
 (d) Neuron 55 RMSE 25.169

Tabel 4. Tingkat Kinerja Model Prediksi PM 2.5 dengan ANN.

Target	Neuron	Lookback	RMSE (μm)	MSE (μm)	MAE (μm)	MAPE (μm)
PM 2.5	5	4	24.673	608.739	21.338	0.194
PM 2.5	10	4	25.185	634.289	21.852	0.196
PM 2.5	15	4	25.138	631.934	21.804	0.195
PM 2.5	20	4	25.221	636.124	21.840	0.196
PM 2.5	25	4	25.190	634.519	21.736	0.194
PM 2.5	30	4	25.226	636.374	21.805	0.195
PM 2.5	35	4	25.181	634.073	21.719	0.194
PM 2.5	40	4	25.164	633.226	21.706	0.194
PM 2.5	45	4	25.188	634.430	21.666	0.193
PM 2.5	50	4	25.155	632.799	21.674	0.193
PM 2.5	55	4	25.114	630.725	21.622	0.193
PM 2.5	60	4	25.093	629.675	21.599	0.193
PM 2.5	65	4	25.198	634.915	21.675	0.193
PM 2.5	70	4	25.132	631.608	21.616	0.193
PM 2.5	75	4	25.150	632.515	21.634	0.193
PM 2.5	80	4	25.167	633.390	21.649	0.193
PM 2.5	85	4	25.140	632.002	21.643	0.193
PM 2.5	90	4	25.134	631.736	21.634	0.193
PM 2.5	95	4	25.258	637.972	21.691	0.194
PM 2.5	100	4	25.111	630.537	21.595	0.193





Gambar 6. Perbandingan Data Aktual dan Prediksi dengan ANN
 (a) Neuron 5 RMSE 24.673 , (b) Neuron 55 RMSE 25.114, (c) Neuron 60 RMSE 25.093,
 (d) Neuron 70 RMSE 25.132

4. Kesimpulan

Penelitian ini dilakukan untuk memprediksi kualitas udara khususnya PM_{2.5} yang dapat menyebabkan gangguan kesehatan pada manusia atau makhluk hidup lainnya. Pada penelitian ini menerapkan dua metode yaitu Long Short-Term Memory dan Artificial Neural Network. Dataset yang digunakan dalam penelitian didapatkan dari Portal Satu Data yang dipublikasikan oleh Dinas Lingkungan Hidup DKI Jakarta. Berdasarkan dataset tersebut digunakan enam atribut untuk memprediksi nilai PM_{2.5}. Untuk nilai korelasi antara atribut dengan target memiliki korelasi koefisien antara 0.01 sampai 0.91. Hasil pengujian dengan metode LSTM menunjukkan hasil yang lebih baik dari ANN dengan nilai RMSE 23.311, MAE 19.391 dan MAPE 0.176. Sedangkan pemodelan menggunakan metode ANN mendapatkan hasil RMSE 24.673, MAE 21.338, MAPE 0.194.

Berdasarkan hasil dari penelitian yang telah didapatkan, model ini dapat diterapkan pada sebuah aplikasi yang mudah diakses baik dalam bentuk mobile, web, maupun desktop. Hal ini tentunya sangat bermanfaat bagi masyarakat karena dapat memberikan peringatan dini terkait kualitas udara yang berpotensi membahayakan kesehatan. Prediksi yang akurat memungkinkan pemerintah untuk mengambil tindakan preventif lebih awal, serta membantu dalam merumuskan kebijakan lingkungan yang lebih efektif dan berbasis data untuk meningkatkan kualitas udara di Jakarta.

Daftar Pustaka

- [1] M. Thoriq Maulana, M. Hilmi Habibullah, Sunandar, N. Sholihah, M. Ainul Rifqi L. P., and F. Fahrudin, "Laporan Akhir: Kegiatan Pemantauan Kualitas Udara Provinsi DKI Jakarta Tahun 2022," *Pemerintah Provinsi Drh. Khusus Ibuk. Jakarta Dinas Lingkung. Hidup*, vol. 1, no. 201310200311137, pp. 78–79, 2022.
- [2] A. Arkadia, B. Hananto, and D. S. Prasvita, "Optimasi Long Short Term Memory Dengan Adam Menggunakan Data Udara Kota DKI Jakarta," *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, pp. 92–101, 2022.
- [3] N. Saurabh, "LSTM - RNN Model to Predict Future Stock Prices using an Efficient Optimizer," *Int. Res. J. Eng. Technol.*, vol. 7, no. 11, pp. 672–677, 2020.
- [4] Y. Jiao, Z. Wang, and Y. Zhang, "Prediction of Air Quality Index Based on LSTM," in *2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, 2019, pp. 17–20. doi: 10.1109/ITAIC.2019.8785602.
- [5] R. Janarthanan, P. Partheeban, K. Somasundaram, and P. Navin Elamparithi, "A deep learning approach for prediction of air quality index in a metropolitan city," *Sustain. Cities Soc.*, vol. 67, p. 102720, 2021, doi: <https://doi.org/10.1016/j.scs.2021.102720>.
- [6] D. L. Hidup, "Data Indeks Standar Pencemar Udara (ISPU) di Provinsi DKI Jakarta," 2022, 2022. <https://satudata.jakarta.go.id> (accessed Jan. 01, 2022).
- [7] I. Nyoman, K. Wardana, N. Jawas, I. Komang, and A. A. Aryanto, "Prediksi Penggunaan Energi Listrik pada Rumah Hunian Menggunakan Long Short-Term Memory," *TIERS Inf. Technol. J.*, vol. 1, no. 1, pp. 1–11, 2020, [Online]. Available: <http://journal.undiknas.ac.id/index.php/tiers>
- [8] I. W. A. Suranata, I. N. K. Wardana, N. Jawas, and I. K. A. A. Aryanto, "Feature engineering and long short-term memory for energy use of appliances prediction," *Telkomnika (Telecommunication Comput. Electron. Control.)*, vol. 19, no. 3, pp. 920–930, 2021, doi: 10.12928/TELKOMNIKA.v19i3.17882.
- [9] B. T. Pham, S. K. Singh, and H. B. Ly, "Using artificial neural network (Ann) for prediction of soil coefficient of consolidation," *Vietnam J. Earth Sci.*, vol. 42, no. 4, pp. 311–319, 2020, doi: 10.15625/0866-7187/42/4/15008.
- [10] Q. T. N. Phan, M. Mondal, and S. Kazushi, "Application of LSTM and ANN Models for Traffic Time

-
- Headway Prediction in Expressway Tollgates,” in *2022 Moratuwa Engineering Research Conference (MERCon)*, 2022, pp. 1–6. doi: 10.1109/MERCon55799.2022.9906226.
- [11] N. He, L. Liu, D. Chu, and C. Qian, “Air Conditioning Cooling Load Prediction Based on LSTM-ANN,” in *2022 5th International Symposium on Autonomous Systems (ISAS)*, 2022, pp. 1–6. doi: 10.1109/ISAS55863.2022.9757345.
- [12] I. N. K. Wardana, J. W. Gardner, and S. A. Fahmy, “Optimising deep learning at the edge for accurate hourly air quality prediction,” *Sensors (Switzerland)*, vol. 21, no. 4, pp. 1–28, 2021, doi: 10.3390/s21041064.
- [13] T. O. Hodson, “Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not,” *Geosci. Model Dev.*, vol. 15, no. 14, pp. 5481–5487, 2022, doi: 10.5194/gmd-15-5481-2022.