

Evaluasi Model Jaringan Saraf Tiruan untuk Prediksi Konsentrasi Nitrogen Oksida (NO_x)

Evaluation of Artificial Neural Network Model for Predicting Nitrogen Oxides (NO_x) Concentration

¹Muhammad Farrih Mahabbataka Arsyada, ²Raras Tyasnurita*

^{1,2}Departemen Sistem Informasi, Fakultas Teknologi Elektro dan Informatika Cerdas,
Institut Teknologi Sepuluh Nopember

^{1,2}Jl. Teknik Kimia, Keputih, Kecamatan Sukolilo, Kota Surabaya, Jawa Timur, Indonesia

*e-mail: raras@is.its.ac.id

(received: 27 June 2024, revised: 9 February 2025, accepted: 17 March 2025)

Abstrak

Nitrogen Oksida (NO_x) adalah salah satu polutan udara yang memerlukan perhatian mendalam karena dapat menimbulkan dampak negatif terhadap kesehatan manusia, lingkungan, dan perekonomian. Penelitian ini penting dilakukan untuk menyediakan model prediksi konsentrasi NO_x yang akurat, yang berguna sebagai dasar dalam mengambil keputusan dan tindakan mitigasi polusi udara secara efektif. Penelitian ini bertujuan untuk mengevaluasi beberapa model jaringan saraf tiruan untuk menentukan model terbaik dalam memprediksi konsentrasi NO_x secara akurat. Salah satu metode yang dapat digunakan untuk memprediksi data polusi udara seperti NO_x adalah jaringan saraf tiruan (neural networks). Pada penelitian ini dibangun empat model ANN (Artificial Neural Networks) untuk melakukan prediksi data konsentrasi NO_x dan dilakukan evaluasi terhadap model yang dapat memberikan performa yang paling optimal. Adapun keempat model tersebut adalah Feed Forward Neural Network (FNN), Time Lagged Neural Networks (TLNN), Seasonal Artificial Neural Networks (SANN), dan Long Short Term Memory (LSTM). Penelitian ini memanfaatkan data konsentrasi NO_x yang berasal dari dataset kualitas udara yang berasal dari UCI Machine Learning Repository. Berdasarkan proses pengujian, menunjukkan hasil bahwa model yang dengan performa terbaik dalam melakukan prediksi dengan nilai eror paling kecil adalah model LSTM dengan rincian parameter 24 input nodes, tiga hidden nodes, satu output nodes, dan 300 epoch. Adapun nilai RMSE untuk LSTM, FNN, TLNN, dan SANN berturut turut adalah 57.3, 62.8, 64, dan 89.

Kata kunci: LSTM, jaringan saraf tiruan, polusi udara, konsentrasi NO_x, model prediktif

Abstract

Nitrogen Oxides (NO_x) are air pollutants that require serious attention due to their potential negative impacts on human health, the environment, and the economy. This research is crucial to provide accurate predictive models of NO_x concentration, which can serve as a foundation for decision-making and effective air pollution mitigation measures. The objective of this study is to evaluate several artificial neural network (ANN) models to determine the most effective model for accurately predicting NO_x concentrations. One of the methods used for predicting air pollution data, such as NO_x, is artificial neural networks (ANN). In this study, four ANN models were constructed and evaluated: Feed Forward Neural Network (FNN), Time Lagged Neural Network (TLNN), Seasonal Artificial Neural Network (SANN), and Long Short-Term Memory (LSTM). The models predict NO_x concentration using data from the air quality dataset provided by the UCI Machine Learning Repository. Testing results indicate that the LSTM model performs best, achieving the lowest error value, characterized by 24 input nodes, three hidden nodes, one output node, and 300 training epochs. The RMSE values for LSTM, FNN, TLNN, and SANN are 57.3, 62.8, 64, and 89, respectively.

Keywords: LSTM, artificial neural networks, air pollution, NO_x concentration, predictive model

1 Pendahuluan

Pencemaran udara menjadi salah satu isu lingkungan yang krusial karena memberikan dampak buruk terhadap kesehatan manusia, lingkungan, dan perekonomian. Menurut laporan IQAir 2023, Indonesia menempati peringkat pertama di Asia Tenggara untuk polusi udara, dengan kota Tangerang Selatan memiliki konsentrasi PM_{2.5} tahunan sebesar 71,7 $\mu\text{g}/\text{m}^3$, delapan kali lipat di atas pedoman WHO [1]. Lebih lanjut, laporan yang dirilis oleh Kementerian Kesehatan RI menyatakan bahwa empat dari sepuluh penyakit dengan kasus terbanyak di Indonesia adalah penyakit pernapasan, seperti PPOK dan pneumonia, yang terkait dengan polusi udara [2]. Nitrogen Oksida (NO_x) adalah salah satu polutan udara yang juga menjadi perhatian utama karena paparan jangka panjang dapat menyebabkan risiko terhadap kesehatan janin, menurunkan tingkat kesuburan, meningkatkan risiko kematian, dan kanker. Konsentrasi tinggi NO_x juga berkontribusi dalam pembentukan ozon troposferik (O₃), suatu oksidan kuat yang dapat menyebabkan fenomena smog atau kabut asap [3].

Beberapa upaya mitigasi yang telah dilakukan meliputi pengendalian emisi kendaraan, peningkatan transportasi publik, penghijauan kota, dan pengendalian emisi industri. Namun demikian, masih terdapat berbagai kendala seperti kurangnya koordinasi antar lembaga, keterbatasan sumber daya, penegakan hukum yang kurang efektif, serta rendahnya partisipasi masyarakat akibat minimnya edukasi dan sosialisasi mengenai pentingnya pengendalian pencemaran udara [4]. Untuk mengatasi berbagai tantangan tersebut, diperlukan pendekatan yang lebih proaktif dalam pengendalian pencemaran udara. Salah satu pendekatan tersebut adalah melalui sistem prediksi kualitas udara. Perkiraan kualitas udara memiliki peran penting dalam pengelolaan kualitas udara, karena dapat memberikan peringatan kesehatan secara tepat waktu, melengkapi program pengendalian emisi yang sudah ada, mendukung perencanaan operasional (misalnya pada sektor penerbangan), serta intervensi darurat untuk mengantisipasi kejadian polusi udara yang tinggi [5].

Salah satu pendekatan yang dapat dimanfaatkan dalam memperkirakan tingkat pencemaran udara adalah dengan menerapkan metode jaringan saraf tiruan (*neural networks*). Jaringan saraf tiruan (ANN) adalah teknik pemodelan data numerik yang bersifat non-linier, terinspirasi dari mekanisme jaringan saraf pada makhluk hidup yang meniru proses kerja otak manusia [6]. Data hasil pemantauan polusi udara merupakan data non-linear dan metode jaringan saraf tiruan telah terbukti efektif dan dapat diandalkan dalam menyelesaikan masalah yang bersifat non-linier [7].

Permasalahan yang akan dibahas dalam penelitian ini adalah bagaimana mengembangkan dan menentukan model jaringan saraf tiruan yang paling akurat untuk memprediksi data polusi udara, seperti konsentrasi NO_x. Hal ini didasarkan oleh fakta bahwa terdapat berbagai model jaringan saraf tiruan dengan akurasi yang cukup tinggi yang dapat digunakan untuk memprediksi data polusi udara [8] [9], tetapi belum banyak penelitian yang membandingkan akurasi dari masing-masing model tersebut. Penelitian ini secara khusus bertujuan untuk mengevaluasi dan membandingkan tingkat akurasi beberapa model ANN tanpa mengembangkan aplikasi atau visualisasi khusus, melainkan berfokus pada identifikasi model prediksi terbaik. Dengan demikian, rekomendasi yang dihasilkan dari penelitian ini akan berupa saran penggunaan model yang optimal sebagai dasar pertimbangan bagi para pengambil kebijakan dalam menentukan metode mitigasi polusi udara yang lebih efektif. Namun, rekomendasi kebijakan secara eksplisit berada di luar ruang lingkup penelitian ini. Hal ini menunjukkan bahwa penelitian ini memiliki potensi untuk memberikan kontribusi yang signifikan terhadap pengembangan model prediksi data polusi udara, terutama dalam konteks pengambilan keputusan oleh pemangku kepentingan terkait pengendalian pencemaran udara.

Dengan demikian, penelitian ini penting dilakukan untuk mengembangkan dan menentukan model jaringan saraf tiruan yang paling akurat untuk memprediksi data polusi udara. Hasil penelitian ini diharapkan dapat digunakan untuk meningkatkan akurasi prediksi polusi udara dan membantu pengendalian polusi udara.

2 Tinjauan Literatur

Berbagai penelitian sebelumnya telah mengkaji performa berbagai model jaringan saraf tiruan (ANN) dalam konteks prediksi polusi udara. Beberapa studi membandingkan berbagai model ANN seperti Feed Forward Neural Network (FNN), Time Lagged Neural Networks (TLNN), Seasonal Artificial Neural Networks (SANN), dan Long Short Term Memory (LSTM). Studi [10] menemukan bahwa pendekatan ANN dengan real-time dynamic error correction mampu menghasilkan prediksi

<http://sistemasi.ftik.unisi.ac.id>

polusi udara yang lebih akurat dibandingkan metode tradisional dalam daerah dengan tingkat polusi tinggi. Studi tersebut, meskipun menunjukkan keunggulan ANN, tidak melakukan perbandingan eksplisit antar jenis ANN yang berbeda.

Penelitian [11] menggunakan data yang diperoleh dari Dinas Lingkungan Hidup Pemerintah Kota Surabaya untuk memprediksi berbagai parameter kualitas udara seperti CO, NO, NO₂, NO_x, PM₁₀, SO₂, dan faktor-faktor meteorologi seperti arah dan kecepatan angin, radiasi global, kelembaban, dan suhu udara. Algoritma *Recurrent Neural Network-Long Short Term Memory* (RNN-LSTM) digunakan dalam melakukan prediksi. Analisis yang dilakukan mencakup analisis tren, korelasi antara polutan dan faktor meteorologi, serta prediksi Karbon Monoksida menggunakan RNN-LSTM di Surabaya yang dikorelasikan dengan data meteorologi. Hasil penelitian menunjukkan bahwa model prediksi terbaik menggunakan RNN-LSTM dengan RMSE sebesar 1,880 dan menggunakan dua *hidden layer* serta 50 *epoch*.

Penelitian [12] memanfaatkan metode deep learning yang termasuk dalam kategori Recurrent Neural Network, seperti Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), dan Gated Recurrent Unit (GRU), dimanfaatkan untuk melakukan prediksi data deret waktu pada kualitas udara berdasarkan parameter suhu, kelembaban, partikel debu halus PM10, serta Indeks Standar Pencemar Udara (ISPU). Hasil penelitian menunjukkan bahwa penerapan model LSTM dan LSTM *Bidirectional* menghasilkan kinerja yang lebih baik daripada model *Gated Recurrent Unit* (GRU). Dalam kedua model tersebut, nilai *Root Mean Square Error* (RMSE) prediksi pada data uji lebih kecil daripada standar deviasi dari dataset uji untuk variabel suhu, kelembaban, dan ISPU dengan skor untuk suhu (LSTM 3.18 < 4.40, LSTM-Bi 3.23 < 4.40), kelembaban (LSTM 6.96 < 19.32, LSTM-Bi 7.07 < 19.32), dan ISPU (LSTM 1.84 < 1.90, LSTM-Bi 1.86 < 1.90).

Penelitian [13] menggunakan metode *long short-term memory* untuk melakukan prediksi terhadap lima parameter zat yang terdapat dalam Indeks Standar Pencemar Udara (ISPU), menghasilkan hasil pengujian yang berbeda untuk setiap parameter tersebut. Partisi data yang digunakan untuk tiap parameter zat berbeda satu sama lain seperti contohnya PM₁₀ dengan 80% data latih dan 20% data uji sedangkan SO₂ dengan 50% data latih dan 50% data uji. Berdasarkan hasil pengujian partisi data, pemodelan terbaik ditemukan untuk PM₁₀ dengan konfigurasi *epoch* = 100, *hidden layer* = 1, jumlah *neuron* = 128, dan *optimizer* = Rmsprop, menghasilkan nilai *Root Mean Square Error* (RMSE) sebesar 0.007237 sedangkan untuk NO₂ menggunakan *epoch* = 50, *hidden layer* = 1, *neuron* = 64, *optimizer* = Adam, menghasilkan nilai RMSE sebesar 0.0431.

Berdasarkan metode *long short-term memory* (LSTM), penelitian [14] menggunakan model prediksi dengan *multi-output and multi-index of supervised learning* (MMSL). Penelitian menggunakan beberapa indikator polusi udara seperti PM2.5, CO, NO₂, O₃, dan SO₂ dan menggunakan indikator evaluasi seperti MAE, RMSE, dan R₂. Model MMSL yang diusulkan, dibandingkan dengan model dasar lainnya, yaitu CNN-LSTM dan CNN-Bi-RNN juga dengan dua model mutakhir yang ada saat ini, yaitu STE dan LSTM-FC. Hasil perbandingan menunjukkan model MMSL lebih efektif daripada model-model lainnya. Hal ini dibuktikan dengan hasil indikator evaluasi, contohnya pada nilai RMSE MMSL, CNN-LSTM, dan CNN-Bi-RNN berturut-turut adalah 11.148, 28.414, dan 28.824.

Penelitian [15] menggunakan analisis korelasi dan *artificial neural networks* (termasuk *wavelet* ANNs atau biasa disingkat WANN) untuk mengidentifikasi hubungan linier dan non-linier berturut-turut pada *air pollution index* (API) dan variabel meteorologi di kota Xi'an dan Lanzhou. *Bayesian regularization* diaplikasikan pada algoritma yang telah dilatih dan menunjukkan bahwa model ANN dan WANN menghasilkan prediksi yang akurat untuk data API di Kota Xi'an dan Lanzhou dengan performa model WANN yang lebih baik dibandingkan dengan ANN, yaitu skor pada model WANN (R = 0.8846 untuk Xi'an dan R = 0.8906 untuk Lanzhou) sedangkan ANN mendapatkan skor (R = 0.8037 untuk Xi'an dan R = 0.7742 untuk Lanzhou). Hasil ini menunjukkan bahwa WANN efektif dalam peramalan API jangka pendek karena dapat mengenali pola historis sehingga WANN dapat mengidentifikasi hubungan non-linier antara variabel input dan output.

Penelitian [16] mengembangkan model untuk memprediksi konsentrasi partikulat (PM) menggunakan metode *long short term memory* (LSTM) dan metode *deep autoencoder* (DAE) kemudian membandingkan hasil model dengan indikator evaluasi *root mean square error* (RMSE). Peneliti menerapkan model tersebut pada data kualitas udara per jam dari 25 stasiun di Seoul, Korea Selatan, untuk periode 1 Januari 2015 hingga 31 Desember 2018. Penelitian memprediksi konsentrasi

<http://sistemasi.ftik.unisi.ac.id>

PM untuk jangka waktu 10 hari dengan laju pembelajaran optimal 0.01 untuk 100 *epochs* dengan *batch size* 32 pada model LSTM dan *batch size* 64 untuk model terbaik model DAE. Model yang diajukan secara efektif dapat memprediksi konsentrasi PM, dengan model LSTM menunjukkan kinerja yang sedikit lebih baik daripada DAE.

Penelitian [17] bertujuan untuk membandingkan performa *multiple linear regression* (MLR) dan *multi-layer perceptron* (MLP) dalam memprediksi konsentrasi SO₂ pada kualitas udara di Kota Tehran. Beberapa parameter berbeda dipilih untuk memprediksi konsentrasi harian SO₂ seperti parameter meteorologi, data lalu lintas perkotaan, informasi ruang terbuka hijau, dan parameter waktu. Berdasarkan hasil yang diperoleh, koefisien korelasi (R_2) dan *root means square error* (RMSE) dari model MLR adalah 0,708 dan 6,025. Sementara itu, nilai RMSE dari model MLP adalah 0,9 dan 0,42. Berdasarkan hasil analisis sensitivitas, nilai *one-day time delay*, indikator taman, musim/tahun, dan luas area taman merupakan faktor utama yang mempengaruhi konsentrasi SO₂. Model MLP yang diusulkan dalam penelitian ini dapat diaplikasikan untuk mendukung, menganalisis, dan meningkatkan prediksi polusi udara dan manajemen kualitas udara.

Penelitian [18] melakukan perbandingan antara model semi-supervised bidirectional LSTM dengan metode ANN lainnya dalam konteks prediksi kualitas udara dan menemukan bahwa model LSTM menunjukkan hasil terbaik karena kemampuannya dalam menangkap pola temporal kompleks. Studi ini menggarisbawahi pentingnya penggunaan pendekatan LSTM untuk data time series kualitas udara yang memiliki pola temporal rumit.

Secara kritis, meskipun model LSTM menunjukkan hasil yang menjanjikan dibandingkan model ANN konvensional lainnya, sebagian besar studi yang ada masih terbatas dalam melakukan evaluasi komparatif secara komprehensif pada berbagai jenis model ANN dalam satu studi yang sama, terutama dalam konteks prediksi konsentrasi NO_x secara spesifik. Oleh karena itu, penelitian ini hadir untuk mengisi kesenjangan tersebut dengan mengevaluasi dan membandingkan kinerja beberapa model ANN dalam konteks prediksi konsentrasi NO_x yang lebih spesifik dan terfokus.

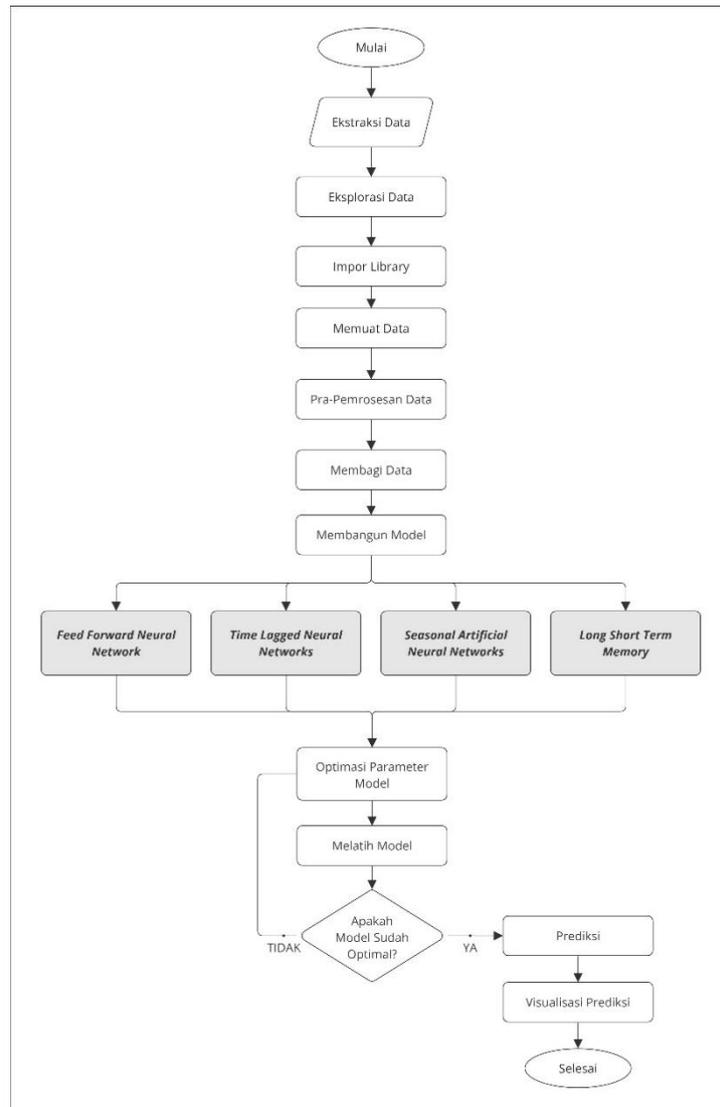
3 Metode Penelitian

Penelitian ini menggunakan pendekatan eksperimental untuk mengevaluasi performa empat model jaringan saraf tiruan dalam memprediksi konsentrasi polutan NO_x. Metodologi yang digunakan mencakup serangkaian tahapan sistematis mulai dari pengumpulan data hingga evaluasi model, dengan mempertimbangkan karakteristik data time series dan kompleksitas masing-masing arsitektur jaringan saraf tiruan. Dalam pelaksanaan proses penelitian, penulis menggunakan bahasa pemrograman Python serta beberapa library pendukung seperti NumPy, Pandas, Keras, Scikit-learn, dan Matplotlib. Tools ini digunakan untuk proses pra-pemrosesan data, pembangunan model, pelatihan, prediksi, serta visualisasi hasil evaluasi model.

Meskipun penelitian ini tidak secara eksplisit membangun aplikasi berbasis pengguna (user interface) atau sistem prediktif terintegrasi, seluruh tahapan pemodelan dan analisis telah dikemas dalam bentuk notebook interaktif (Jupyter Notebook) yang memungkinkan pengembangan lebih lanjut ke arah implementasi sistem atau aplikasi. Visualisasi hasil seperti grafik evaluasi dan perbandingan prediksi model disusun dalam bentuk yang mudah dipahami.

Struktur tahapan penelitian ini secara tidak langsung mengadopsi prinsip-prinsip dalam framework CRISP-DM (Cross Industry Standard Process for Data Mining), mulai dari tahap business understanding (pemahaman masalah), data understanding (eksplorasi dataset), data preparation (pra-pemrosesan), modeling (pembangunan dan pelatihan model), hingga evaluation (penilaian performa model). Penelitian ini belum mencakup tahap deployment secara eksplisit, namun struktur tahapan memungkinkan perluasan ke arah pengembangan sistem prediktif terapan di masa mendatang.

Tahapan penelitian secara lengkap dapat dilihat pada Gambar 1, yang mengilustrasikan alur kerja mulai dari ekstraksi data mentah, pra-pemrosesan, pembangunan model, hingga analisis performa menggunakan berbagai metrik evaluasi. Setiap tahapan dirancang untuk memastikan validitas hasil dan memungkinkan perbandingan yang objektif antar model.



Gambar 1. Tahapan penelitian

Pada tahap awal, dataset diunduh dari tautan <https://archive.ics.uci.edu/dataset/360/air+quality> [19]. Pemilihan dataset ini didasarkan pada beberapa pertimbangan. Dataset ini memiliki karakteristik multivariate time-series, yang sangat sesuai dengan pendekatan pemodelan yang digunakan dalam penelitian ini, yaitu jaringan saraf tiruan (ANN). Selain itu, dataset ini menyajikan data yang direkam secara real-time dari lingkungan nyata, sehingga meningkatkan validitas hasil pemodelan terhadap fenomena sebenarnya di lapangan. Kemudian, salah satu variabel utama yang tersedia dalam dataset ini adalah konsentrasi Nitrogen Oksida (NO_x), yang menjadi fokus utama dalam penelitian ini. Keberadaan variabel NO_x memungkinkan peneliti untuk melakukan prediksi spesifik terhadap konsentrasi polutan tersebut.

Dataset ini dipublikasikan pada 22 Maret 2016 dan memuat data rata-rata respons setiap jam dari lima sensor kimia berbasis oksida logam yang terintegrasi dalam sistem Multisensor untuk pemantauan kualitas udara. Perangkat ini diletakkan di lingkungan terbuka pada wilayah yang memiliki tingkat polusi tinggi, tepatnya di sisi jalan di salah satu kota di Italia. Setiap respons per jam tersebut dicatat bersamaan dengan data referensi konsentrasi gas dari alat analisis bersertifikasi. Dataset air quality memiliki karakteristik multivariate time series, yaitu kumpulan deret waktu yang menggambarkan berbagai aspek dari fenomena temporal tertentu [20]. Record yang terdapat pada dataset merupakan pengukuran riil yang dilakukan di dunia nyata, dengan total 9358 baris data dan 15 kolom pengamatan.

Record pada dataset direkam dari Maret 2004 hingga Februari 2005 (satu tahun) yang menjadi rekaman terpanjang yang tersedia secara bebas dari respon perangkat sensor kimia kualitas udara yang digunakan di lapangan. Konsentrasi rata-rata rata-rata per jam untuk CO, Hidrokarbon Non Metana, Benzena, Total Nitrogen Oksida (NO_x) dan Nitrogen Dioksida (NO₂), dihasilkan oleh *reference certified analyzer* yang berada di lokasi yang sama. Dataset ini dapat digunakan secara eksklusif untuk tujuan penelitian. Namun, tidak diperbolehkan untuk keperluan komersial. Dataset tersebut juga tidak memiliki nilai null yang berarti dataset memiliki kualitas yang cukup baik.

Dataset memiliki total 15 kolom dengan kolom pertama dan kedua berturut-turut adalah tanggal dan waktu pengamatan. Dataset ini mengukur kualitas udara pada rentang waktu setiap jam sehingga untuk pengamatan selama satu hari terdapat total 24 pengamatan dimana setiap pengamatan mewakili pengamatan setiap jam. Sedangkan kolom-kolom yang lain mengukur aneka rata-rata konsentrasi dari berbagai macam gas polusi udara, juga rata-rata respon sensor per jamnya. Untuk penelitian ini, dipilih salah satu dari variabel, yaitu NO_x(GT) yang merupakan konsentrasi NO_x rata-rata per jam yang sebenarnya dalam ppb dengan jumlah 569 baris pertama data.

Setelah melakukan ekstraksi data NO_x dari dataset mentah, selanjutnya dilakukan eksplorasi data untuk mengetahui pola dan karakteristik data. Bahasa pemrograman yang digunakan pada penelitian ini adalah Python sehingga perlu mengimpor beberapa *library* yang dibutuhkan pada proses prediksi nantinya. Data hasil ekstraksi sebelumnya dimuat ke dalam sebuah variabel kemudian disesuaikan formatnya menjadi matriks dan dinormalisasi. Data dibagi menjadi *training set* dan *test set* dengan jumlah persentase untuk setiap set berturut-turut adalah 80% dan 20%. Selain pembagian data tersebut, validasi model juga menjadi perhatian dalam proses evaluasi. Namun dalam penelitian ini belum diterapkan teknik cross-validation seperti k-fold cross-validation secara eksplisit. Teknik ini menjadi salah satu peluang pengembangan untuk penelitian selanjutnya, agar mampu memberikan gambaran generalisasi model yang lebih menyeluruh terhadap dataset yang tersedia.

Selanjutnya adalah membuat setiap model yang akan digunakan untuk memprediksi data NO_x, yaitu *Feed Forward Neural Network* (FNN), *Time Lagged Neural Networks* (TLNN), *Seasonal Artificial Neural Networks* (SANN), dan *Long Short Term Memory* (LSTM). Setiap model tersebut akan menghasilkan nilai prediksi untuk jangka waktu yang ditentukan. Kemudian setiap model akan dievaluasi menggunakan beberapa metrik statistik yang biasa digunakan untuk mengukur seberapa akurat model memprediksi nilai sebenarnya seperti *mean squarred error* (MSE), *mean absolute error* (MAE), dan nilai *root mean square error* (RMSE). Nilai ketiga metrik tersebut nantinya akan digunakan sebagai acuan untuk menentukan model artificial neural networks yang memberikan performa terbaik dengan nilai eror paling kecil. Penelitian ini bertujuan untuk melakukan evaluasi terhadap 4 model ANN yang dibangun untuk memprediksi konsentrasi NO_x sehingga dihasilkan model yang memberikan performa paling optimal.

4 Hasil dan Pembahasan

Penelitian ini melakukan prediksi data konsentrasi polutan udara NO_x menggunakan beberapa model, yaitu *Feed Forward Neural Network* (FNN), *Time Lagged Neural Networks* (TLNN), *Seasonal Artificial Neural Networks* (SANN), dan *Long Short Term Memory* (LSTM). Kemudian hasil prediksi akan dievaluasi menggunakan beberapa indikator evaluasi seperti MSE, MAE, dan RMSE untuk menentukan model terbaik yang menghasilkan nilai prediksi paling akurat dengan nilai eror yang paling kecil. Penelitian ini mengimplementasikan beberapa *library* berbeda antara lain adalah *numpy*, *pandas*, *matplotlib*, *keras*, *sklearn*, dan *IPython*. Keempat model yang akan digunakan untuk melakukan prediksi data NO_x dilatih dengan bahasa pemrograman Python.

Hal pertama yang dilakukan adalah mengunduh *raw dataset* dari laman <https://archive.ics.uci.edu/dataset/360/air+quality> [19]. Preview dataset dalam format csv dapat dilihat pada Tabel 1. Dataset memiliki 15 kolom atau variabel pengamatan dengan dua kolom pertama merupakan tanggal dan waktu pengukuran. Variabel NO_x diketahui terdapat pada kolom ke-8. Selanjutnya perlu dilakukan ekstraksi variabel NO_x beserta tanggal dan waktu pengukurannya.

Tabel 2 menunjukkan hasil ekstraksi yang menampilkan tiga variabel pengukuran, yaitu konsentrasi NO_x, tanggal, dan waktu pengukuran. Pada penelitian ini hanya akan menggunakan 569 data pertama dari total jumlah keseluruhan dataset. Proses tersebut menghasilkan dataset dengan 3

variabel pengukuran, yaitu konsentrasi NO_x, tanggal, dan waktu pengukuran dengan total *record* sebanyak 569 baris.

Tabel 1. Dataset air quality

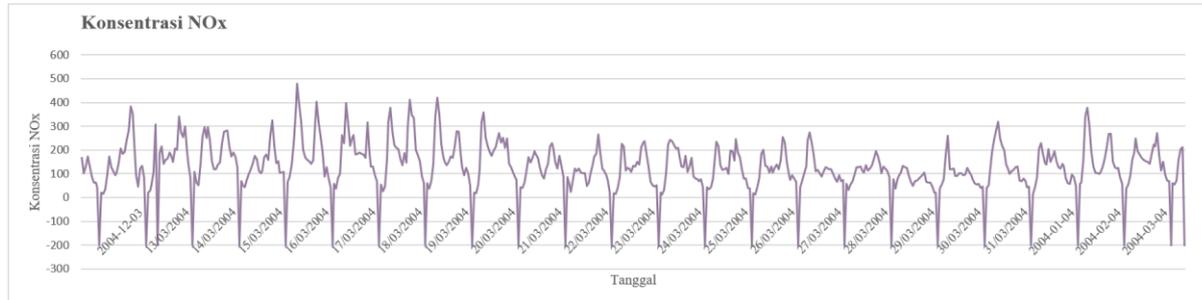
Date	Time	CO(GT)	PT08.S1(CO)	NMHC(GT)	C6H6(GT)	PT08.S2(NMHC)	NOx(GT)	PT08.S3(NOx)	NO2(GT)	PT08.S4(NO2)	PT08.S5(O3)	T	RH	AH
10/3/2004	18.00.00	2,6		1360	150 11,9		1046	166	1056	113	1692	1268 13,6	48,9	0,7578
10/3/2004	19.00.00	2		1292	112 9,4		955	103	1174	92	1559	972 13,3	47,7	0,7255
10/3/2004	20.00.00	2,2		1402	88 9,0		939	131	1140	114	1555	1074 11,9	54,0	0,7502
10/3/2004	21.00.00	2,2		1376	80 9,2		948	172	1092	122	1584	1203 11,0	60,0	0,7867
10/3/2004	22.00.00	1,6		1272	51 6,5		836	131	1205	116	1490	1110 11,2	59,6	0,7888
10/3/2004	23.00.00	1,2		1197	38 4,7		750	89	1337	96	1393	949 11,2	59,2	0,7848
11/3/2004	00.00.00	1,2		1185	31 3,6		690	62	1462	77	1333	733 11,3	56,8	0,7603
11/3/2004	01.00.00	1		1136	31 3,3		672	62	1453	76	1333	730 10,7	60,0	0,7702
11/3/2004	02.00.00	0,9		1094	24 2,3		609	45	1579	60	1276	620 10,7	59,7	0,7648
11/3/2004	03.00.00	0,6		1010	19 1,7		561	-200	1705	-200	1235	501 10,3	60,2	0,7517
11/3/2004	04.00.00	-200		1011	14 1,3		527	21	1818	34	1197	445 10,1	60,5	0,7465
11/3/2004	05.00.00	0,7		1066	8 1,1		512	16	1918	28	1182	422 11,0	56,2	0,7366
11/3/2004	06.00.00	0,7		1052	16 1,6		553	34	1738	48	1221	472 10,5	58,1	0,7353
11/3/2004	07.00.00	1,1		1144	29 3,2		667	98	1490	82	1339	730 10,2	59,6	0,7417
11/3/2004	08.00.00	2		1333	64 8,0		900	174	1136	112	1517	1102 10,8	57,4	0,7408
11/3/2004	09.00.00	2,2		1351	87 9,5		960	129	1079	101	1583	1028 10,5	60,6	0,7691
11/3/2004	10.00.00	1,7		1233	77 6,3		827	112	1218	98	1446	860 10,8	58,4	0,7552
11/3/2004	11.00.00	1,5		1179	43 5,0		762	95	1328	92	1362	671 10,5	57,9	0,7352
11/3/2004	12.00.00	1,6		1236	61 5,2		774	104	1301	95	1401	664 9,5	66,8	0,7951
11/3/2004	13.00.00	1,9		1286	63 7,3		869	146	1162	112	1537	799 8,3	76,4	0,8393
11/3/2004	14.00.00	2,9		1371	164 11,5		1034	207	983	128	1730	1037 8,0	81,1	0,8736
11/3/2004	15.00.00	2,2		1310	79 8,8		933	184	1082	126	1647	946 8,3	79,8	0,8778
.....
.....

Tabel 2. Hasil ekstraksi data no_x

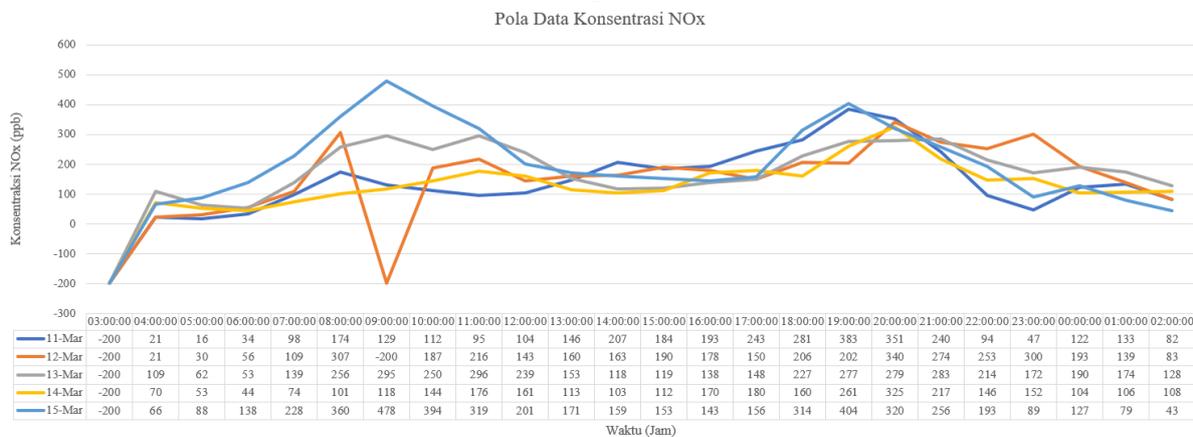
Nomor	Tanggal	Waktu	NO _x (GT)
1	2004-10-03	18:00:00	166
2	2004-10-03	19:00:00	103
3	2004-10-03	20:00:00	131
4	2004-10-03	21:00:00	172
5	2004-10-03	22:00:00	131
...
...
...
565	2004-03-04	06:00:00	73
566	2004-03-04	07:00:00	159
567	2004-03-04	08:00:00	203
568	2004-03-04	09:00:00	211
569	2004-03-04	10:00:00	-200

Pada eksplorasi data, hal utama yang dilakukan adalah mengidentifikasi pola data *time series*. Hal tersebut penting dilakukan sebelum melakukan *forecasting* karena beberapa alasan, diantaranya supaya mendapatkan pemahaman yang lebih baik terkait perilaku data, meningkatkan akurasi, meminimalisir kesalahan prediksi, meningkatkan efisiensi, dan meningkatkan keandalan prediksi. Mengidentifikasi pola data *time series* dapat dilakukan dengan berbagai cara, salah satunya yaitu dengan membuat visualisasi data menggunakan *line graph*. Gambar 2 menunjukkan visualisasi data konsentrasi NO_x yang disajikan dalam bentuk *line graph*. Pada Gambar 2 terlihat bahwa data mempunyai pola yang berulang setiap beberapa periode waktu. Setiap periode waktu tertentu jumlah konsentrasi NO_x akan turun pada titik minimumnya kemudian naik dengan fluktuatif hingga kemudian turun lagi pada titik minimumnya. Apabila dilakukan pengamatan terhadap pola data, diketahui bahwa konsentrasi NO_x akan mencapai titik minimumnya pada nilai -200 ppb yang terjadi setiap jam 03.00 setiap hari, seperti ditunjukkan pada Gambar 3. Berdasarkan Gambar 3 dapat dikatakan bahwa pola tersebut berulang setiap periode 24 jam atau satu hari. Kemudian pada Gambar 2 diketahui

bahwa pola yang berulang pada grafik tidak memiliki tren positif maupun negatif. Selain itu, panjang pola pada grafik tidak berubah seiring dengan perkembangan waktu. Berdasarkan informasi-informasi tersebut dapat disimpulkan bahwa data mempunyai pola *constant trend with additive seasonality*. Informasi ini berguna supaya penulis dapat lebih baik dalam mengetahui perilaku data dan meningkatkan akurasi saat proses *forecasting* nantinya.

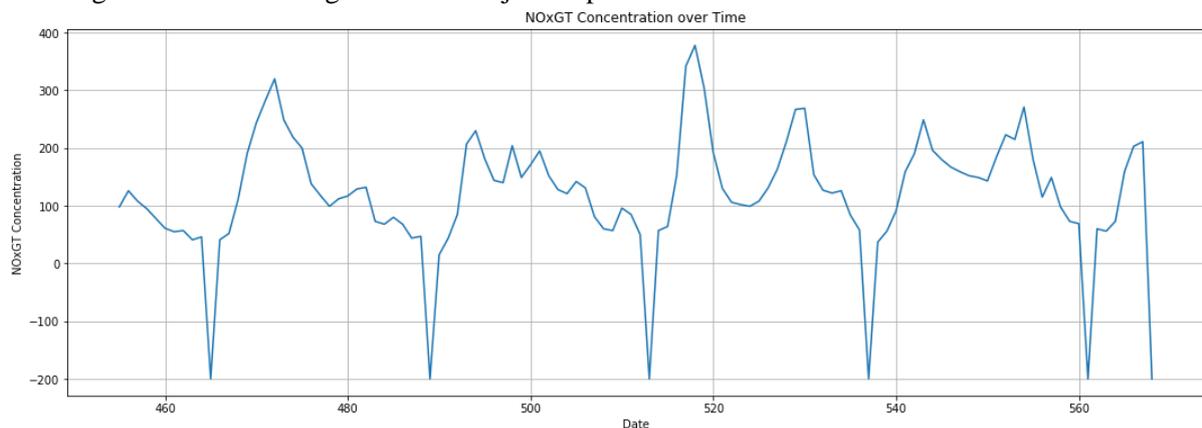


Gambar 2. Grafik garis konsentrasi no_x



Gambar 3. Pola data konsentrasi no_x

Sebelum memuat data yang dilakukan pada bahasa pemrograman Python, dilakukan proses impor library seperti numpy, pandas, matplotlib, keras, sklearn, dan IPython. Selanjutnya pada pra-pemrosesan data, data diubah menjadi format yang kompatibel dengan ANN, yaitu dalam bentuk matriks dan dinormalisasi menggunakan MinMaxScaler. Kemudian dataset yang telah dikonversi menjadi bentuk matriks akan dibagi ke dalam dua macam data, yaitu training set dan testing set dengan komposisi berturut-turut adalah 80% dan 20%. Data training dimanfaatkan dalam proses pelatihan model, sedangkan data testing berperan dalam mengukur performa model [21]. Gambar 4 menunjukkan data *testing set* dari konsentrasi NO_x dalam bentuk visualisasi *line graph* yang nantinya akan digunakan untuk mengevaluasi kinerja setiap model.



Gambar 4. Grafik data uji

Secara umum, terdapat 4 fungsi yang diperlukan dalam membangun algoritma keempat model ANN yang diajukan, yaitu pra-premrosesan data, prakiraan langkah waktu 'n' berikutnya, proses pelatihan dan prediksi, dan menghitung skor pada data pengujian. Keempat fungsi tersebut akan digunakan pada proses pelatihan model nantinya. Setiap model memiliki parameter yang nantinya akan digunakan dalam proses pelatihan dan prediksi sesuai cara kerja setiap model tersebut. Pada FNN, parameter yang dibutuhkan adalah *look_back*, *hidden_nodes*, *output_nodes*, *epochs*, *batch_size*, dan *future_steps*. Pada model TLNN memiliki parameter *time_lagged_points*, *hidden_nodes*, *output_nodes*, *epochs*, *batch_size*, dan *future_steps*. Pada model SANN memiliki parameter *seasonal_period*, *hidden_nodes*, *epochs*, *batch_size*, dan *future_steps*. Kemudian pada model LSTM memiliki parameter *look_back*, *hidden_nodes*, *output_nodes*, *epochs*, *batch_size*, dan *future_steps*. Pemilihan nilai parameter pada setiap model menentukan nilai hasil prediksi data NO_x sehingga diperlukan beberapa kali percobaan supaya mendapatkan model yang optimal. Apabila model masih belum optimal maka diperlukan penyesuaian kembali nilai-nilai parameter yang dimasukkan pada model. Nilai-nilai parameter tersebut juga harus disesuaikan dengan karakteristik data yang telah diidentifikasi pada tahap eksplorasi data sebelumnya supaya mendapatkan hasil yang optimal.

Tabel 3. Nilai parameter terbaik setiap model

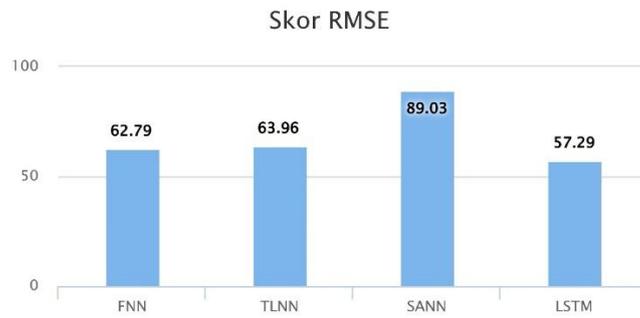
Model	Input Nodes	Hidden Nodes	Output Nodes	Epoch
FNN	25	6	1	500
TLNN	12	4	1	300
SANN	24	4	1	500
LSTM	24	3	1	300

Tabel 3 menunjukkan kombinasi nilai parameter terbaik untuk masing-masing model dimana pada pengujian yang dilakukan dengan 20 *batch size* dan $n = 60$ untuk prakiraan langkah waktu 'n' berikutnya (*number of future steps forecasted*) didapatkan bahwa untuk model FNN parameter terbaiknya adalah 25 *input nodes*, 6 *hidden nodes*, 1 *output nodes*, dan 500 *epoch*. Parameter terbaik untuk TLNN adalah 12 *time lagged points*, 4 *hidden nodes*, 1 *output nodes*, dan 300 *epoch*. Parameter terbaik untuk SANN adalah 24 *input nodes*, 4 *hidden nodes*, 24 *output nodes*, dan 500 *epoch*. Parameter terbaik untuk LSTM adalah 24 *input nodes*, 3 *hidden nodes*, 1 *output nodes*, dan 300 *epoch*. Hasil prediksi menunjukkan bahwa model LSTM adalah model yang memberikan nilai prediksi paling akurat dengan hasil prediksi untuk setiap model ditampilkan pada Gambar 6.

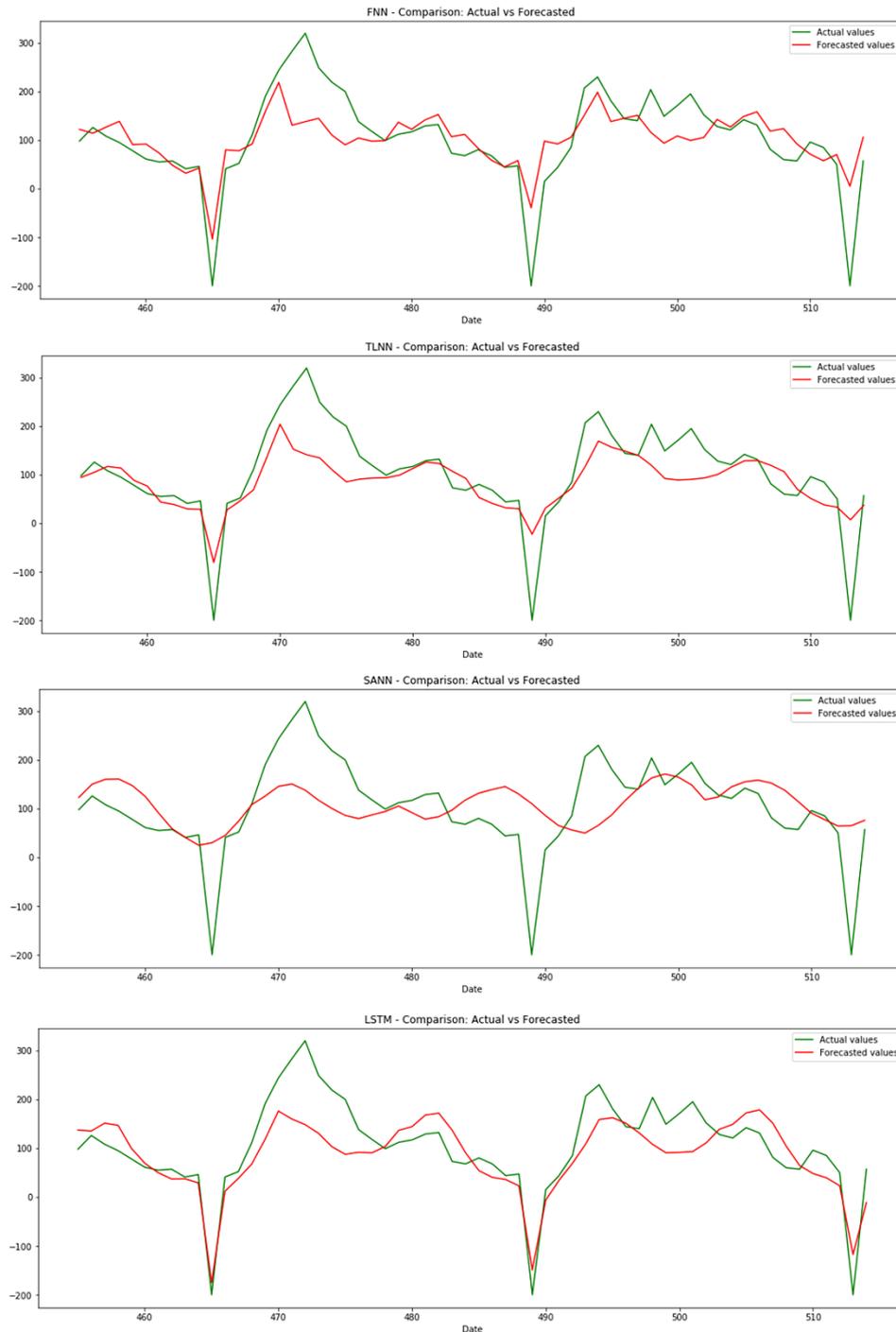
Tabel 4. Nilai metrik evaluasi setiap model

Model	MSE	MAE	RMSE
FNN	3942.3	43.6	62.8
TLNN	4091.4	42.9	64.0
SANN	7926.1	61.3	89.0
LSTM	3281.9	44.7	57.3

Tabel 4 menunjukkan nilai metrik evaluasi untuk masing-masing model. Berdasarkan hasil nilai indikator *mean squarred error* (MSE), *mean absolute error* (MAE), dan *root mean squarred error* (RMSE) diketahui bahwa model LSTM adalah model yang mempunyai nilai eror paling kecil di antara ketiga model lainnya dengan MSE, MAE, dan RMSE pada model LSTM berturut-turut adalah 3281.9, 44.7, dan 57.3. Pada model FNN nilai MSE, MAE, dan RMSE berturut-turut adalah 3942.3, 43.6, dan 62.8. Pada model TLNN nilai MSE, MAE, dan RMSE berturut-turut adalah 4091.4, 42.9, dan 64. Pada model SANN nilai MSE, MAE, dan RMSE berturut-turut adalah 7926.1, 61.3, dan 89. Adapun nilai RMSE untuk keempat model divisualisasikan dalam bentuk *bar chart* pada Gambar 5.



Gambar 5. Grafik perbandingan nilai rmse keempat model



Gambar 6. Grafik hasil prediksi NO_x keempat model ANN

Berdasarkan hasil prediksi dan beberapa nilai indikator evaluasi, diketahui bahwa model LSTM merupakan model ANN yang memiliki performa terbaik dalam melakukan prediksi data *time series* NO_x dibandingkan dengan ketiga model ANN lainnya, yaitu FNN, TLNN, dan SANN. Model LSTM berhasil melakukan prediksi data NO_x dengan rincian parameter 24 *input nodes*, 3 *hidden nodes*, 1 *output nodes*, 300 epoch serta nilai MSE, MAE, dan RMSE berturut-turut adalah 3281.9, 44.7, dan 57.3. Hal ini konsisten dengan penelitian-penelitian terkait sebelumnya dimana LSTM mendapatkan performa terbaik dalam melakukan prediksi dibanding dengan model-model ANN lainnya [11] [13] [16].

Pada Gambar 6, terlihat bahwa model FNN cukup sensitif terhadap kenaikan dan penurunan konsentrasi NO_x . Prediksi yang dihasilkan mengikuti tren data aktual dengan baik, meskipun terdapat beberapa deviasi pada puncak dan lembah. Hal ini menunjukkan bahwa FNN memiliki kemampuan yang cukup baik dalam menangkap pola perubahan konsentrasi NO_x , namun tidak selalu akurat pada perubahan yang sangat cepat atau ekstrem.

Model TLNN juga menunjukkan sensitivitas yang baik terhadap perubahan konsentrasi NO_x . Grafik prediksi TLNN memperlihatkan bahwa model ini mampu mengikuti fluktuasi data aktual dengan cukup baik. Namun, ada beberapa perbedaan yang signifikan pada titik-titik perubahan yang tajam, yang menunjukkan bahwa TLNN mungkin memiliki keterbatasan dalam menangkap perubahan mendadak dalam data musiman.

Grafik prediksi SANN menunjukkan performa yang lebih rendah dibandingkan model lainnya. Model ini tampaknya tidak mampu menangkap fluktuasi dan pola musiman dengan baik, yang terlihat dari perbedaan besar antara prediksi dan data aktual. Meskipun data memiliki pola musiman, performa rendah SANN mungkin disebabkan oleh kompleksitas tambahan yang dihasilkan oleh faktor musiman, yang tidak diakomodasi dengan baik oleh model ini.

Model LSTM menunjukkan hasil prediksi yang paling akurat dibandingkan model lainnya. Grafik prediksi LSTM memperlihatkan bahwa model ini sangat sensitif terhadap kenaikan dan penurunan konsentrasi NO_x , dengan prediksi yang sangat mendekati data aktual. Kemampuan LSTM untuk mengingat informasi jangka panjang memungkinkan model ini menangkap pola kompleks dan perubahan musiman dengan lebih baik.

Hasil ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa LSTM unggul dalam memprediksi data *time-series* dengan pola musiman dan tren yang dinamis. Misalnya, penelitian [18] juga menemukan bahwa model LSTM memberikan performa lebih baik dibandingkan dengan model ANN konvensional dalam konteks prediksi kualitas udara. Namun, beberapa penelitian lain menunjukkan bahwa dalam kondisi tertentu, model seperti TLNN atau SANN dapat memiliki performa yang lebih stabil dalam dataset dengan pola musiman yang sangat kuat. Perbedaan ini mungkin disebabkan oleh karakteristik dataset yang digunakan, di mana pola NO_x dalam dataset ini lebih dipengaruhi oleh faktor jangka panjang dibandingkan pola musiman yang berulang secara spesifik.

Kontribusi utama penelitian ini adalah memberikan evaluasi komprehensif terhadap berbagai model ANN dalam konteks prediksi konsentrasi NO_x , yang masih jarang dikaji secara spesifik dalam studi sebelumnya. Dengan hasil yang menunjukkan keunggulan LSTM, penelitian ini dapat menjadi referensi bagi penelitian lanjutan dalam pengembangan model prediktif polusi udara yang lebih akurat. Dari perspektif praktis, model yang lebih akurat dalam memprediksi polusi udara seperti LSTM dapat digunakan dalam berbagai kebijakan pengendalian polusi. Misalnya, prediksi polusi yang lebih tepat dapat membantu otoritas lingkungan dalam merancang strategi mitigasi yang lebih responsif, seperti pembatasan lalu lintas kendaraan di area tertentu atau peningkatan efisiensi sistem peringatan dini bagi masyarakat yang rentan terhadap paparan polusi udara. Oleh karena itu, hasil penelitian ini memiliki potensi besar untuk diimplementasikan dalam sistem pemantauan dan perencanaan lingkungan yang lebih cerdas.

5 Kesimpulan

Berdasarkan evaluasi terhadap empat model jaringan saraf tiruan (ANN) yang digunakan untuk memprediksi konsentrasi NO_x , yaitu FNN, TLNN, SANN, dan LSTM, ditemukan bahwa model LSTM memberikan performa terbaik dengan nilai error paling rendah (MSE: 3281.9, MAE: 44.7,

RMSE: 57.3) menggunakan parameter optimal 24 input nodes, 3 hidden nodes, 1 output node, dan 300 epoch. Model FNN menempati posisi kedua (RMSE: 62.8), diikuti TLNN (RMSE: 64.0), sementara SANN menunjukkan performa terendah (RMSE: 89.0) dalam memprediksi pola musiman data. Keunggulan LSTM terutama terletak pada kemampuannya menangkap pola jangka panjang dan mengatasi masalah vanishing gradient serta exploding gradient. Untuk pengembangan lebih lanjut, penelitian ini merekomendasikan eksplorasi dataset yang lebih besar dan lebih bervariasi untuk meningkatkan generalisasi model terhadap kondisi lingkungan yang berbeda. Selain itu, teknik peningkatan model seperti fine-tuning hyperparameter yang lebih luas atau penerapan pendekatan hybrid ANN, seperti kombinasi LSTM dengan metode ensemble atau attention mechanism, dapat dipertimbangkan untuk meningkatkan akurasi prediksi. Dari perspektif aplikasi praktis, model prediksi NO_x yang lebih akurat dapat dimanfaatkan dalam sistem pemantauan kualitas udara berbasis IoT atau diterapkan dalam sistem prediksi berbasis cloud yang memberikan informasi real-time kepada pembuat kebijakan dan masyarakat. Implementasi ini dapat membantu otoritas lingkungan dalam menyusun strategi mitigasi yang lebih tepat, seperti pembatasan lalu lintas di area dengan tingkat polusi tinggi atau peringatan dini bagi kelompok rentan terhadap paparan polusi udara. Oleh karena itu, hasil penelitian ini diharapkan dapat berkontribusi pada pengembangan solusi berbasis AI dalam pengendalian kualitas udara yang lebih efektif dan adaptif.

Referensi

- [1] G. Indonesia, "Laporan Kualitas Udara Dunia IQAir 2023: Indonesia Terburuk se- Asia Tenggara," *greenpeace.org*, Mar. 19, 2024. <https://www.greenpeace.org/indonesia/siaran-pers-2/58036/laporan-kualitas-udara-dunia-iqair-2023-indonesia-terburuk-se-asia-tenggara/> (accessed Mar. 14, 2025).
- [2] Rokom, "Polusi Udara Sebabkan Angka Penyakit Respirasi Tinggi," *kemkes.go.id*, Apr. 04, 2023. <https://sehatnegeriku.kemkes.go.id/baca/trilis-media/20230404/2642721/polusi-udara-sebabkan-angka-penyakit-respirasi-tinggi/> (accessed Mar. 14, 2025).
- [3] H. Haruna, L. Lahming, F. Amir, dan A. R. Asrib, "Pencemaran Udara Akibat Gas Buang Kendaraan Bermotor dan Dampaknya terhadap Kesehatan," *UNM Environmental Journals*, vol. 2, no. 2, p. 57, Apr. 2019, doi: <https://doi.org/10.26858/uej.v2i2.10092>.
- [4] G. M. Maula, "Efektivitas Implementasi Kebijakan Pengendalian Pencemaran Udara di Indonesia," *Savana: Indonesian Journal of Natural Resources and Environmental Law*, vol. 1, no. 2, pp. 145–159, Aug. 2024, doi: <https://doi.org/10.25134/savana.v1i2.223>.
- [5] NOAA, *Air Quality Forecasting: A Review of Federal Programs and Research Needs*. Boulder, CO: National Oceanic and Atmospheric Administration, 2001. [Online]. Available: <https://www.esrl.noaa.gov/csd/aqrsd/reports/forecasting.pdf> (accessed: Mar. 15, 2025).
- [6] L. S. Rini, O. T. Karya, dan F. Sirait, "Implementasi Jaringan Syaraf Tiruan pada Pendeteksi Keaslian dan Nominal Uang," *Jurnal Teknologi Elektro*, vol. 12, no. 2, p. 65, Jul. 2021, doi: <https://doi.org/10.22441/jte.2021.v12i2.004>.
- [7] J. Veri, S. Surmayanti, dan G. Guslendra, "Prediksi Harga Minyak Mentah menggunakan Jaringan Syaraf Tiruan," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 3, pp. 503–512, Jul. 2022, doi: <https://doi.org/10.30812/matrik.v21i3.1382>.
- [8] S. Agarwal, S. Sharma, S. R., M. H. Rahman, S. Vranckx, B. Maiheu, L. Blyth, S. Janssen, P. Gargava, V. K. Shukla, S. Batra, "Air Quality Forecasting using Artificial Neural Networks with Real Time Dynamic Error Correction in Highly Polluted Regions," *Science of the Total Environment*, vol. 735, no. 735, pp. 139454–139454, Sep. 2020, doi: <https://doi.org/10.1016/j.scitotenv.2020.139454>.
- [9] L. Zhang, P. Liu, L. Zhao, G. Wang, W. Zhang, dan J. Liu, "Air Quality Predictions with a Semi-Supervised Bidirectional LSTM Neural Network," *Atmospheric Pollution Research*, vol. 12, no. 1, pp. 328–339, Jan. 2021, doi: <https://doi.org/10.1016/j.apr.2020.09.003>.
- [10] S. Agarwa, S. Sharma, R. Suresh, Md. H. Rahman, S. Vranckx, B. Maiheu, L. Blyth, S. Janssen, P. Gargava, V. K. Shukla, S. D. Batra, "Air Quality Forecasting using Artificial Neural Networks with Real Time Dynamic Error Correction in Highly Polluted Regions," *Science of the Total*

- Environment, vol. 735, pp. 139454–139454, Sep. 2020, doi: <https://doi.org/10.1016/j.scitotenv.2020.139454>.
- [11] M. A. Faishol, “Analisis Data Runtun Waktu Prediksi Polusi Udara di Kota Surabaya menggunakan Deep Learning RNN-LSTM,” Thesis, Sepuluh Nopember Institute of Technology, 2020. Accessed: Jan. 31, 2024. [Online]. Available: <http://repository.its.ac.id/id/eprint/76659>
- [12] Y. Karyadi, “Prediksi Kualitas Udara dengan Metoda LSTM, Bidirectional LSTM, dan GRU,” *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 1, pp. 671–684, Mar. 2022, doi: <https://doi.org/10.35957/jatisi.v9i1.1588>.
- [13] B. K. Hidayatullah, M. Kallista, and C. Setianingsih, “Prediksi Indeks Standar Pencemaran Udara menggunakan Metode Long Short-Term Memory berbasis Web (Studi Kasus pada Kota Jakarta),” in *e-Proceeding of Engineering*, Telkom University, Jun. 2022, pp. 1274–1255.
- [14] D. Seng, Q. Zhang, X. Zhang, G. Chen, and X. Chen, “Spatiotemporal Prediction of Air Quality Based on LSTM Neural Network,” *Alexandria Engineering Journal*, vol. 60, no. 2, pp. 2021–2032, Apr. 2021, doi: <https://doi.org/10.1016/j.aej.2020.12.009>.
- [15] Q. Guo, Z. He, S. Li, X. Li, J. Meng, Z. Hou, J. Liu, Y. Chen, “Air Pollution Forecasting using Artificial and Wavelet Neural Networks with Meteorological Conditions,” *Aerosol and Air Quality Research*, vol. 20, no. 6, pp. 1429–1439, 2020, doi: <https://doi.org/10.4209/aaqr.2020.03.0097>.
- [16] T. Xayasouk, H. Lee, and G. Lee, “Air Pollution Prediction using Long Short-Term Memory (LSTM) and Deep Autoencoder (DAE) Models,” *Sustainability*, vol. 12, no. 6, p. 2570, Mar. 2020, doi: <https://doi.org/10.3390/su12062570>.
- [17] S. R. Shams, A. Jahani, S. Kalantary, M. Moeinaddini, and N. Khorasani, “The Evaluation on Artificial Neural Networks (ANN) and Multiple Linear Regressions (MLR) Models for Predicting SO₂ Concentration,” *Urban Climate*, vol. 37, p. 100837, May 2021, doi: <https://doi.org/10.1016/j.uclim.2021.100837>.
- [18] L. Zhang, P. Liu, L. Zhao, G. Wang, W. Zhang, and J. Liu, “Air Quality Predictions with a Semi-Supervised Bidirectional LSTM Neural Network,” *Atmospheric Pollution Research*, vol. 12, no. 1, pp. 328–339, Jan. 2021, doi: <https://doi.org/10.1016/j.apr.2020.09.003>.
- [19] S. Vito, “Air Quality,” *UCI Machine Learning Repository*, 2016. [Online]. Available: <https://doi.org/10.24432/C59K5F>.
- [20] J. Li, H. Izakian, W. Pedrycz, and I. Jamal, “Clustering-based Anomaly Detection in Multivariate Time Series Data,” *Applied Soft Computing*, vol. 100, p. 106919, Mar. 2021, doi: <https://doi.org/10.1016/j.asoc.2020.106919>.
- [21] D. Musfiroh, U. Khaira, P. E. P. Utomo, and T. Suratno, “Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset menggunakan InSet Lexicon,” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 1, no. 1, pp. 24–33, Mar. 2021, doi: <https://doi.org/10.57152/malcom.v1i1.20>.