



SKRIPSI

PENERAPAN ALGORITMA CNN-LSTM DENGAN LAGGED FEATURES UNTUK PREDIKSI KONSENTRASI ZAT POLUTAN UDARA DI KOTA YOGYAKARTA

AFIFA SALSABILA

NPM 21081010130

DOSEN PEMBIMBING

Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.

Achmad Junaidi, S.Kom., M.Kom.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025**



SKRIPSI

PENERAPAN ALGORITMA CNN-LSTM DENGAN LAGGED FEATURES UNTUK PREDIKSI KONSENTRASI ZAT POLUTAN UDARA DI KOTA YOGYAKARTA

AFIFA SALSABILA

NPM 21081010130

DOSEN PEMBIMBING

Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.

Achmad Junaidi, S.Kom., M.Kom.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025**

LEMBAR PENGESAHAN

PENERAPAN ALGORITMA CNN-LSTM DENGAN LAGGED FEATURES UNTUK PREDIKSI KONSENTRASI ZAT POLUTAN UDARA DI KOTA YOGYAKARTA

Oleh :

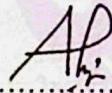
AFIFA SALSABILA

NPM. 21081010130

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 04 Juni 2025.

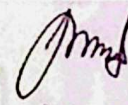
Menyetujui

Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.
NPT. 222198 60 816400


.....

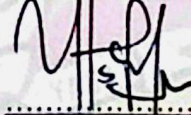
(Pembimbing I)

Achmad Junaidi, S.Kom., M.Kom.
NPT. 3 7811 04 0199 1


.....

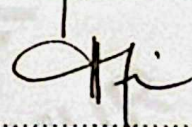
(Pembimbing II)

Yisti Vita Via, S.ST., M.Kom.
NIP. 19860425 202121 2 001


.....

(Ketua Penguji)

Afina Lina Nurlaili, S.Kom., M.Kom.
NIP. 1993121 3202203 2 010


.....

(Penguji I)

Mengetahui,
Dekan Fakultas Ilmu Komputer


Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

**PENERAPAN ALGORITMA CNN-LSTM DENGAN LAGGED
FEATURES UNTUK PREDIKSI KONSENTRASI ZAT POLUTAN
UDARA DI KOTA YOGYAKARTA**

Oleh:

AFIFA SALSABILA

NPM. 21081010130

Menyetujui,

**Koordinator Program Studi Informatika
Fakultas Ilmu Komputer**



Fetty Tri Anggraeny, S.Kom., M.Kom.
NIP. 19820211 202121 2 005

SURAT PERNYATAAN BEBAS PLAGIASI

Yang bertanda tangan di bawah ini:

Nama Mahasiswa : Afifa Salsabila
NPM : 21081010130
Program : Sarjana (S1)
Program Studi : Informatika
Fakultas : Fakultas Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila dikemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan dipergunakan sebagaimana mestinya.



Surabaya, 18 Juni 2025

Yang Membuat Pernyataan,

Afifa Salsabila
NPM. 21081010130

ABSTRAK

Nama Mahasiswa / NPM : Afifa Salsabila / 21081010130
Judul Skripsi : Penerapan Algoritma CNN-LSTM dengan Lagged Features untuk Prediksi Konsentrasi Zat Polutan Udara di Kota Yogyakarta
Dosen Pembimbing : 1. Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.
2. Achmad Junaidi, S.Kom., M.Kom.

Penelitian ini dilaksanakan untuk mengetahui kinerja algoritma *Convolutional Neural Network* (CNN) - *Long Short Term Memory* (LSTM) dalam melakukan prediksi konsentrasi zat polutan udara. Dalam penelitian ini, terdapat beberapa tahapan penelitian, yaitu pengumpulan dan pembersihan data, *preprocessing* data, penerapan teknik *lagged features*, pembuatan model CNN-LSTM, pelatihan model, evaluasi model, serta pengujian model. Data dalam penelitian ini merupakan data konsentrasi karbon monoksida (CO) dan data meteorologi harian Kota Yogyakarta. Setelah diproses, data akan menjadi masukan bagi model untuk kemudian dapat melakukan proses prediksi. Pembangunan model CNN-LSTM disesuaikan dengan kondisi data yang ada melalui adanya sistem pengujian. Hasil penelitian memperlihatkan bahwa dalam kasus ini, model CNN-LSTM terbaik didapatkan dengan beberapa pengaturan, seperti pembagian data dengan komposisi 70% *data training* dan 30% *data testing*; 3 *layer* CNN dengan unit neuron masing-masing sebanyak 64; 2 *layer* LSTM dengan unit neuron masing-masing 32; nilai parameter *alpha* pada optimasi Ridge Regression adalah 1.0; serta *learning rate* ada pada angka 0,0005. Melalui pengaturan-pengaturan tersebut, didapatkan hasil metrik evaluasi MSE sebesar 0.00039, RMSE dengan nilai 0.0197, MAE sebesar 0.0152, serta MAPE ada pada angka 6.88%.

Kata kunci : *Convolutional Neural Network* (CNN), *Lagged Features*, *Long Short Term Memory* (LSTM), Kualitas Udara, Prediksi.

ABSTRACT

Student Name / NPM : Afifa Salsabila / 21081010130
Thesis Title : Implementation of CNN-LSTM Algorithm with Lagged Features for Predicting Air Pollutant Concentration in Yogyakarta City
Advisor : 1. Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.
2. Achmad Junaidi, S.Kom., M.Kom.

ABSTRACT

This study aims to evaluate the performance of the Convolutional Neural Network (CNN) - Long Short Term Memory (LSTM) algorithm in predicting air pollutant concentration levels. The research process includes data collection and cleaning, data preprocessing, the application of lagged feature techniques, the creation and training of the CNN-LSTM model, as well as model evaluation and testing. The data used consists of carbon monoxide (CO) concentration and daily meteorological data from Yogyakarta City. After processing, the data is used as input for the prediction model. The development of the CNN-LSTM model is tailored to the data conditions through a testing system. The results show that the best CNN-LSTM model was achieved with the following settings: a data split of 70% for training and 30% for testing; 3 CNN layers with 64 neurons per layer; 2 LSTM layers with 32 neurons per layer; an alpha value of 1.0 for Ridge Regression optimization; and a learning rate of 0.0005. With these settings, the evaluation metrics obtained were: MSE of 0.00039, MAE of 0.0152, RMSE of 0.0197, and MAPE of 6.88%.

Keywords: *Air Quality, Convolutional Neural Network (CNN), Lagged Features, Long Short Term Memory (LSTM), Prediction.*

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul **“Penerapan Algoritma CNN-LSTM dengan *Lagged Features* untuk Prediksi Konsentrasi Zat Polutan Udara di Kota Yogyakarta”** dapat terselesaikan dengan baik. Skripsi ini merupakan bagian dari pencapaian penulis dalam menyelesaikan pendidikan pada program studi Informatika di Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Penulis menyadari bahwa penyusunan skripsi ini tidak lepas dari bantuan, dukungan, dan bimbingan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Ibu Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT. dan Bapak Achmad Junaidi, S.Kom., M.Kom., selaku pembimbing skripsi yang telah memberikan arahan, bimbingan, dan motivasi selama proses penyusunan skripsi ini.
2. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Fetty Tri Anggraeny, S.Kom. M.Kom., selaku Ketua Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Ibu Yisti Vita Via, S.ST., M.Kom. dan Ibu Afina Lina Nurlaili, S.Kom., M.Kom., selaku dosen penguji yang telah memberikan masukan dan kritik yang membangun dalam rangka penyempurnaan skripsi ini.
5. Kepada orangtua penulis, Siti Asiyah dan Cahyo Wijayanto, yang senantiasa mendukung penulis dalam menyelesaikan pendidikan. Penulis mempersembahkan skripsi dan gelar ini kepada keduanya. Terima kasih karena telah percaya pada penulis dan berperan besar dalam menggapai apa yang sudah dicita-citakan.

6. Kepada saudara dan keluarga besar penulis. Terima kasih kepada Arya Arsyavin, Avriela Shanza, serta Pipit Yulia yang telah mendukung dan menjadi penyemangat penulis dalam menyelesaikan skripsi.
7. Kepada sahabat-sahabat penulis, Andhini Putri, Rizky Amelia, dan Najwa Laila, yang telah menemani, menjadi pendengar, sekaligus memberikan saran kepada penulis.
8. Kepada teman-teman penulis semasa kuliah. Terima kasih untuk Mutiq, Nesya, Rayya, Hafiyah, Irsyad, Raihan, Agil, Hubed, dll., untuk semua *support* yang telah diberikan selama penulis menyelesaikan penelitian ini.
9. Kepada sahabat-sahabat penulis semasa sekolah. Terima kasih Bunga Restu, Michelle Cindy, Raka Agung, Nandhita Putri, serta Meylisa Maulidia, yang telah menjadi motivasi penulis untuk menyelesaikan skripsi.
10. Semua pihak yang tidak dapat penulis sebutkan satu per satu, yang telah memberikan kontribusi baik langsung maupun tidak langsung dalam penyelesaian skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna dan banyak kekurangan. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi kesempurnaan penelitian ini di masa mendatang. Semoga skripsi ini dapat memberikan manfaat bagi perkembangan ilmu pengetahuan, bagi para pembaca, serta bagi penulis sendiri.

Surabaya, 18 Juni 2025

Penulis,

Afifa Salsabila

DAFTAR ISI

LEMBAR PENGESAHAN	ii
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI.....	iv
ABSTRAK	v
KATA PENGANTAR.....	vii
DAFTAR GAMBAR.....	xi
DAFTAR TABEL	xv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan Penelitian	4
1.4. Manfaat Penelitian	4
1.5. Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA.....	7
2.1 Penelitian Terdahulu	7
2.2 Kualitas Udara.....	10
2.3 Polusi Udara	11
2.4 ISPU	12
2.5 Data Preprocessing.....	13
2.6 Rolling Mean.....	14
2.7 Forward Fill.....	15
2.8 Backward Fill	16
2.9 Lagged Features	18
2.10 Convolutional Neural Network (CNN).....	19
2.11 Long Short-Term Memory (LSTM)	20
2.12 Model CNN-LSTM.....	22
2.13 Evaluasi Model.....	23
BAB III DESAIN DAN IMPLEMENTASI SISTEM	29
3.1 Metode Penelitian.....	29
3.2 Studi Literatur	30
3.3 Pengumpulan Data	30

3.4 Pre-processing Data.....	33
3.5 Penerapan Teknik Lagged-Features	37
3.6 Pembagian Data.....	38
3.7 Pengembangan Model CNN-LSTM.....	38
3.8 Evaluasi Model.....	39
3.9 Skenario Pengujian.....	40
BAB IV PENGUJIAN DAN ANALISIS.....	43
4.1 Metode Pengujian.....	43
4.1.1 Pengumpulan Data.....	43
4.1.2 Data <i>Pre-processing</i>	44
4.1.3 Model CNN-LSTM	54
4.1.4 Optimasi Model CNN-LSTM.....	58
4.1.5 Evaluasi Model CNN-LSTM.....	59
4.2 Hasil Pengujian.....	60
4.2.1 Skenario Uji Pembagian Data.....	60
4.2.2 Skenario Uji Jumlah <i>Layer</i> CNN.....	69
4.2.3 Skenario Uji Jumlah <i>Layer</i> LSTM	77
4.2.4 Skenario Uji Unit Neuron CNN	87
4.2.5 Skenario Uji Unit Neuron LSTM	95
4.2.6 Skenario Uji Parameter <i>Alpha</i> Ridge Regression.....	103
4.2.7 Hasil Pengujian Data Uji	112
4.3 Evaluasi Model.....	114
4.3.1 Model Prediksi menggunakan Algoritma CNN	114
4.3.2 Model Prediksi menggunakan Algoritma LSTM.....	116
4.3.3 Hasil Evaluasi Model.....	119
4.3.4 Pengembangan <i>Website</i>	120
BAB V PENUTUP	129
5.1 Kesimpulan.....	129
5.2 Saran Pengembangan.....	130
DAFTAR PUSTAKA	131

DAFTAR GAMBAR

Gambar 2. 1 Diagram Arsitektur <i>Layer</i> CNN.....	19
Gambar 2. 2 Diagram Arsitektur <i>Layer</i> LSTM.....	21
Gambar 2. 3 Diagram Arsitektur Model CNN-LSTM.....	23
Gambar 3. 1 Diagram Alur Penelitian	29
Gambar 3. 2 Data Konsentrasi Polutan Udara Harian Kota Yogyakarta.....	31
Gambar 3. 3 Data Meteorologis Harian Kota Yogyakarta.....	32
Gambar 3. 4 Diagram Alur Proses Handling Missing <i>Values</i>	36
Gambar 3. 5 Diagram Alur Proses Normalisasi Data	37
Gambar 4. 1 Data setelah Penanganan <i>Missing Values</i>	46
Gambar 4. 2 Data setelah penambahan <i>Lag Features</i>	48
Gambar 4. 3 Data setelah penambahan <i>Rolling Features</i>	48
Gambar 4. 4 Data setelah Penghapusan Missing Values	49
Gambar 4. 5 Data setelah Proses Normalisasi	51
Gambar 4. 6 Hasil Metrik Evaluasi Pengujian Pembagian Data 60:40	61
Gambar 4. 7 Visualisasi Training loss dan Validation loss Pengujian Pembagian Data 60:40	62
Gambar 4. 8 Visualisasi Actual vs Predicted CO Values Pengujian Pembagian Data 60:40.....	63
Gambar 4. 9 Hasil Metrik Evaluasi Pengujian Pembagian Data 70:30	64
Gambar 4. 10 Visualisasi Training loss dan Validation loss Pengujian Pembagian Data 70:30	65
Gambar 4. 11 Visualisasi Actual vs Predicted CO Values Pengujian Pembagian Data 70:30	66
Gambar 4. 12 Hasil Metrik Evaluasi Pengujian Pembagian Data 80:20	67
Gambar 4. 13 Visualisasi Training loss dan Validation loss Pengujian Pembagian Data 80:20	67
Gambar 4. 14 Visualisasi Actual vs Predicted CO Values Pengujian Pembagian Data 80:20	68
Gambar 4. 15 Hasil Metrik Evaluasi Pengujian 2 Layer CNN	70

Gambar 4. 16 Visualisasi Training loss dan Validation loss Pengujian 2 Layer CNN	70
Gambar 4. 17 Visualisasi Actual vs Predicted CO Values Pengujian 2 Layer CNN	71
Gambar 4. 18 Hasil Metrik Evaluasi Pengujian 3 Layer CNN	72
Gambar 4. 19 Visualisasi Training loss dan Validation loss Pengujian 3 Layer CNN	73
Gambar 4. 20 Visualisasi Actual vs Predicted CO Values Pengujian 3 Layer CNN	74
Gambar 4. 21 Hasil Metrik Evaluasi Pengujian 4 Layer CNN	75
Gambar 4. 22 Visualisasi Training loss dan Validation loss Pengujian 4 Layer CNN	75
Gambar 4. 23 Visualisasi Actual vs Predicted CO Values Pengujian 4 Layer CNN	76
Gambar 4. 24 Hasil Metrik Evaluasi Pengujian 2 Layer LSTM	79
Gambar 4. 25 Visualisasi Training loss dan Validation loss Pengujian 2 Layer LSTM	80
Gambar 4. 26 Visualisasi Actual vs Predicted CO Values Pengujian 2 Layer LSTM	81
Gambar 4. 27 Hasil Metrik Evaluasi Pengujian 3 Layer LSTM	81
Gambar 4. 28 Visualisasi Training loss dan Validation loss Pengujian 3 Layer LSTM	82
Gambar 4. 29 Visualisasi Actual vs Predicted CO Values Pengujian 3 Layer LSTM	83
Gambar 4. 30 Hasil Metrik Evaluasi Pengujian 4 Layer LSTM	84
Gambar 4. 31 Visualisasi Training loss dan Validation loss Pengujian 4 Layer LSTM	84
Gambar 4. 32 Visualisasi Actual vs Predicted CO Values Pengujian 4 Layer LSTM	85
Gambar 4. 33 Hasil Metrik Evaluasi Pengujian 16 Neuron CNN.....	88
Gambar 4. 34 Visualisasi Training loss dan Validation loss Pengujian 16 Neuron CNN.....	88

Gambar 4. 35 Visualisasi Actual vs Predicted CO Values Pengujian 16 Neuron CNN	89
Gambar 4. 36 Hasil Metrik Evaluasi Pengujian 32 Neuron CNN	90
Gambar 4. 37 Visualisasi Training loss dan Validation loss Pengujian 32 Neuron CNN	90
Gambar 4. 38 Visualisasi Actual vs Predicted CO Values Pengujian 32 Neuron CNN	91
Gambar 4. 39 Hasil Metrik Evaluasi Pengujian 64 Neuron CNN	92
Gambar 4. 40 Visualisasi Training loss dan Validation loss Pengujian 64 Neuron CNN	93
Gambar 4. 41 Visualisasi Actual vs Predicted CO Values Pengujian 64 Neuron CNN	94
Gambar 4. 42 Hasil Metrik Evaluasi Pengujian 16 Neuron LSTM	96
Gambar 4. 43 Visualisasi Training loss dan Validation loss Pengujian 16 Neuron LSTM	97
Gambar 4. 44 Visualisasi Actual vs Predicted CO Values Pengujian 16 Neuron LSTM	98
Gambar 4. 45 Hasil Metrik Evaluasi Pengujian 32 Neuron LSTM	98
Gambar 4. 46 Visualisasi Training loss dan Validation loss Pengujian 32 Neuron LSTM	99
Gambar 4. 47 Visualisasi Actual vs Predicted CO Values Pengujian 32 Neuron LSTM	100
Gambar 4. 48 Hasil Metrik Evaluasi Pengujian 64 Neuron LSTM	101
Gambar 4. 49 Visualisasi Training loss dan Validation loss Pengujian 64 Neuron LSTM	101
Gambar 4. 50 Visualisasi Actual vs Predicted CO Values Pengujian 64 Neuron LSTM	102
Gambar 4. 51 Hasil Metrik Evaluasi Pengujian Paramter Alpha 1.0.....	104
Gambar 4. 52 Visualisasi Training loss dan Validation loss Pengujian Parameter Alpha 1.0.....	105
Gambar 4. 53 Visualisasi Actual vs Predicted CO Values Pengujian Parameter Alpha 1.0.....	106

Gambar 4. 54 Hasil Metrik Evaluasi Pengujian Paramter Alpha 2.0	107
Gambar 4. 55 Visualisasi Training loss dan Validation loss Pengujian Parameter Alpha 2.0	107
Gambar 4. 56 Visualisasi Actual vs Predicted CO Values Pengujian Parameter Alpha 2.0	108
Gambar 4. 57 Hasil Metrik Evaluasi Pengujian Paramter Alpha 3.0	109
Gambar 4. 58 Visualisasi Training loss dan Validation loss Pengujian Parameter Alpha 3.0	110
Gambar 4. 59 Visualisasi Actual vs Predicted CO Values Pengujian Parameter Alpha 3.0	111
Gambar 4. 60 Hasil Pengujian Data Uji dengan Nilai Aktual dan Nilai Prediksi	113
Gambar 4. 61 Hasil Metrik Evaluasi Model CNN	115
Gambar 4. 62 Visualisasi Training loss dan Validation loss Model CNN	115
Gambar 4. 63 Visualisasi Actual vs Predicted CO Values Model CNN	116
Gambar 4. 64 Hasil Metrik Evaluasi Model LSTM	117
Gambar 4. 65 Visualisasi Training loss dan Validation loss Model LSTM	118
Gambar 4. 66 Visualisasi Actual vs Predicted CO Values Model LSTM	119
Gambar 4. 67 Halaman Utama <i>Website</i> Prediksi	125
Gambar 4. 68 Halaman Pemilihan Tanggal oleh Pengguna	127
Gambar 4. 69 Halaman <i>Website</i> Hasil Prediksi	128

DAFTAR TABEL

Tabel 3. 1 Data Konsentrasi Zat CO	30
Tabel 3. 2 Data Meteorologi Harian	32
Tabel 3. 3 Data setelah Penerapan Metode Lagged Features.....	37
Tabel 3. 4 Skenario Pengujian Model CNN-LSTM.....	41
Tabel 4. 1 Hasil Skenario Pengujian Pembagian Data.....	69
Tabel 4. 2 Hasil Skenario Pengujian Jumlah Layer CNN.....	77
Tabel 4. 3 Hasil Pengujian Jumlah Layer LSTM.....	86
Tabel 4. 4 Hasil Pengujian Unit Neuron CNN.....	95
Tabel 4. 5 Hasil Pengujian Unit Neuron LSTM	102
Tabel 4. 6 Hasil Pengujian Nilai Parameter Alpha	111
Tabel 4. 7 Hasil Evaluasi Model	119