

**PERBANDINGAN MODEL PERAMALAN KUALITAS UDARA
MENGUNAKAN METODE LSTM DAN HYBRID ARIMA - LSTM
(STUDI KASUS : PROVINSI SUMATERA SELATAN)**

(Skripsi)

Oleh

NABILA FIRZARIANI

2115061065



**FAKULTAS TEKNIK
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG**

2025

**PERBANDINGAN MODEL PERAMALAN KUALITAS UDARA
MENGUNAKAN METODE LSTM DAN HYBRID ARIMA - LSTM
(STUDI KASUS : PROVINSI SUMATERA SELATAN)**

Oleh

NABILA FIRZARIANI

Skripsi

**Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA TEKNIK**

Pada

**Jurusan Teknik Elektro
Fakultas Teknik Universitas Lampung**



**FAKULTAS TEKNIK
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG**

2025

ABSTRAK

PERBANDINGAN MODEL PERAMALAN KUALITAS UDARA MENGUNAKAN METODE LSTM DAN HYBRID ARIMA – LSTM (STUDI KASUS : PROVINSI SUMATERA SELATAN)

Oleh

NABILA FIRZARIANI

Kualitas udara yang buruk akan memberikan dampak negatif bagi kesehatan masyarakat maupun lingkungan. Indeks Standar Pencemaran Udara (ISPU) merupakan standar yang digunakan lembaga pemerintah Indonesia untuk mengukur tingkat polusi di suatu wilayah. Berdasarkan data ISPU, kualitas udara di Provinsi Sumatera Selatan khususnya kota Palembang berada pada kondisi tidak sehat yang dapat menyebabkan berbagai macam penyakit. Pemantauan kualitas udara menjadi sebuah langkah strategis sebagai upaya mencegah peningkatan pencemaran udara, salah satunya dengan memanfaatkan *machine learning* untuk melakukan peramalan (*forecasting*). Penelitian ini bertujuan untuk membangun, menganalisis serta membandingkan kinerja model *Long Short – Short Term Memory* (LSTM) dan *hybrid* ARIMA – LSTM dalam memprediksi kualitas udara berdasarkan data polutan yang diperoleh dari Dinas Pertanahan dan Lingkungan Hidup Provinsi Sumatera Selatan dari tahun 2020 hingga 2024. *Preprocessing* data dilakukan dengan mengimputasi *missing value* menggunakan nilai mean dan median. Kemudian, model dievaluasi menggunakan metrik *Root Mean Squared Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE) untuk menentukan model terbaik dalam melakukan peramalan. Hasil penelitian menunjukkan bahwa model yang dibangun menggunakan data yang diimputasi dengan nilai median memberikan performa terbaik. Secara keseluruhan model LSTM menunjukkan kinerja yang lebih baik (RMSE = 5,4746; MAPE = 4,04%) dibandingkan dengan model *hybrid* ARIMA – LSTM (RMSE = 5,7482; MAPE = 4,89%).

Kata kunci : *Hybrid* ARIMA – LSTM, ISPU, Kualitas Udara, LSTM, *Machine Learning*

ABSTRACT

COMPARATIVE STUDY OF AIR QUALITY FORECASTING MODELS USING LSTM AND HYBRID ARIMA–LSTM (CASE STUDY: SOUTH SUMATRA PROVINCE)

By

NABILA FIRZARIANI

Poor air quality can negatively impact both public health and the environment. The Air Pollution Standard Index (ISPU) is a benchmark used by the Indonesian government to assess pollution levels in specific regions. According to ISPU data, air quality in South Sumatra Province, particularly in Palembang City, is classified as unhealthy and poses various health risks. Monitoring air quality has become a strategic effort to prevent further deterioration, one of which involves leveraging machine learning for forecasting. This study aims to develop, analyze, and compare the performance of Long Short-Term Memory (LSTM) and hybrid ARIMA–LSTM models in predicting air quality, using pollutant data obtained from the South Sumatra Provincial Office of Land and Environment from 2020 to 2024. The data preprocessing stage involved imputing missing values using both mean and median methods. The models were evaluated using Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) to identify the most accurate forecasting model. The results show that the model trained on data imputed using the median method yielded the best performance. Overall, the LSTM model achieved higher accuracy (RMSE = 5.4746; MAPE = 4.04%) compared to the hybrid ARIMA–LSTM model (RMSE = 5.7482; MAPE = 4.89%).

Keywords: Air Quality, Hybrid ARIMA–LSTM, ISPU, LSTM, Machine Learning

Judul Skripsi

: PERBANDINGAN MODEL PERAMALAN
KUALITAS UDARA MENGGUNAKAN
METODE LSTM DAN HYBRID ARIMA -
LSTM (STUDI KASUS : PROVINSI
SUMATERA SELATAN)

Nama Mahasiswa

: Nabila Firzariani

No. induk Pokok Mahasiswa

: 2115061065

Program Studi

: Teknik Informatika

Jurusan

: Teknik Elektro

Fakultas

: Teknik

MENYETUJUI

1. Komisi Pembimbing

Pembimbing Utama

Pembimbing Pendamping

Ir. Muhamad Komarudin, S. T., M. T.

NIP. 196812071997031006

Ir. Titin Yulianti, S.T., M. Eng.

NIP. 1988070920190320015

2. Mengetahui

Ketua Jurusan

Teknik Elektro

Ketua Program Studi

Teknik Informatika

Herlinawati, S.T., M.T.

NIP. 197103141999032001

Yessi Mulyani, S. T., M.T.

NIP. 197312262000122001

MENGESAHKAN

1. Tim Penguji

Ketua : Ir. Muhamad Komarudin, S. T., M.T.

Sekretaris : Ir. Titin Yulianti, S.T., M. Eng.

Penguji : Yessi Mulyani, S.T., M.T.

2. Dekan Fakultas Teknik

Dr. Eng. Ir. Helmy Fitriawan, S.T., M. Sc.

NIP. 197509282001121002

Tanggal Lulus Ujian Skripsi : 10 Juli 2025

SURAT PERNYATAAN

Dengan ini saya menyatakan bahwa skripsi berjudul "Perbandingan Model Peramalan Kualitas Udara Menggunakan Metode Lstm Dan Hybrid Arima – Lstm (Studi Kasus : Provinsi Sumatera Selatan)" sepenuhnya merupakan hasil karya saya sendiri. Apabila di kemudian hari terbukti pernyataan ini tidak benar, saya siap menerima sanksi sesuai dengan ketentuan hukum yang berlaku.

Bandar Lampung, 10 Juli 2025

Penulis



Nabila Firzariani

NPM. 2115061065

RIWAYAT HIDUP



Penulis dilahirkan di Kota Pagar Alam pada tanggal 06 Juli 2003 sebagai anak pertama dari tiga bersaudara, dari pasangan Bapak Firzal Hendri dan Ibu Tiara Fitriani. Penulis menyelesaikan jenjang Pendidikan formal di SDS Muhammadiyah 1 Pagar Alam pada tahun 2015, kemudian melanjutkan Pendidikan menengah pertama di SMPN 1 Kota Pagar Alam dan lulus pada tahun 2018, serta menamatkan Pendidikan menengah atas di SMAN 1 Kota Pagar Alam pada tahun 2021. Pada tahun 2021, penulis terdaftar sebagai mahasiswa Program Studi Teknik Informatika Universitas Lampung melalui jalur SBMPTN. Selama menjadi mahasiswa, penulis aktif berpartisipasi dalam berbagai kegiatan, antara lain :

1. Menjadi anggota Divisi Media Informasi Himpunan Mahasiswa Teknik Elektro Universitas Lampung periode 2022
2. Menjadi anggota Divisi Penelitian dan Pengembangan Himpunan Mahasiswa Teknik Elektro Universitas Lampung periode 2023
3. Mengikuti kegiatan Studi Independen Bersertifikat dari Kementerian Pendidikan dan Budaya di mitra Hacktivate Teknologi Indonesia pada tahun 2023
4. Menjadi Asisten Laboratorium Teknik Komputer Universitas Lampung tahun akademik 2023/2024
5. Melaksanakan Kuliah Kerja Nyata di Desa Rantau Temiang, Kecamatan Banjit, Kabupaten Way Kanan, Provinsi Lampung pada bulan Januari sampai Februari 2024
6. Mengikuti kegiatan Magang Bersertifikat dari Kementerian Pendidikan dan Budaya di mitra Sekretariat Jenderal DPR RI sebagai Data Engineer pada tahun 2024

MOTTO

”Maka, sesungguhnya berserta kesulitan ada kemudahan.
Sesungguhnya berserta kesulitan ada kemudahan”

(Q.S. Al – Insyirah: 5 – 6)

”Dunia cenderung menuju titik kesimbangan, termasuk apapun yang ada di
dalamnya. Maka, seimbangilah harapanmu dengan upaya.
Rasamu dengan doa. Kelabumu dengan warna.”

(J.S. Khairen, *Melangkah*)

”Kita tidak tahu apa yang disimpan masa depan untuk seseorang. Jangan pernah
meremehkan siapa pun. Selama langit masih di atas dan tanah masih bisa diinjak,
Apapun bisa terjadi.”

(J.S. Khairen, *Kami Bukan Sarjana Kertas*)

PERSEMBAHAN

Puji syukur saya panjatkan atas kehadiran Allah SWT. Yang telah melimpahkan segala rahmat, hidayah dan kekuatan yang diberikan selama proses penyusunan skripsi ini. Karya ini saya persembahkan sebagai bentuk rasa syukur atas segala perjuangan, doa dan dukungan dari orang – orang tercinta yang senantiasa hadir dalam setiap langkah dalam perjuangan ini.

Dengan penuh cinta dan rasa syukur yang tak terhingga, kupersembahkan skripsi ini kepada kedua orang tuaku tercinta, Ayah Firzal dan Ibu Tiara yang selalu memberikan doa, semangat dan dukungan dalam setiap langkah hidupku. Terima kasih atas segala kasih sayang dan pengorbanan yang tak pernah lelah kalian berikan untuk putri kecil kalian ini. Semoga karya sederhana ini dapat menjadi wujud bakti dan rasa syukur atas kasih sayang dan cinta kalian yang tak ternilai.

Teruntuk saudaraku tersayang, Farhan dan Farisa yang menjadi sumber semangat, tawa dan bahagia di saat lelah. Terima kasih atas kerja sama, bantuan dan kebersamaan dalam perjalanan yang telah kita lalui bersama. Semoga dengan segala kekurangan dan kelebihan yang ada, kehadiranku dapat menjadi motivasi kalian untuk terus tumbuh, belajar serta menjadi pribadi yang lebih baik lagi.

Teruntuk diri sendiri, yang telah berjuang sekuat tenaga menghadapi tantangan hingga dapat menyelesaikan skripsi ini. Terima kasih sudah mau berjuang sejauh ini, melewati masa – masa sulit dan rumit. Terima kasih tetap bertahan dan tidak menyerah pada hal – hal yang membuatmu merasa kecil. Semoga semua impian dan harapan yang selama ini kamu perjuangkan dapat tercapai satu per satu. Teruslah melangkah, tumbuh dan belajar untuk menjadi versi terbaik dari dirimu sendiri.

SANWACANA

Bismillahirramanirrahim, puji Syukur selalu terpanjatkan kepada Allah SWT. atas segala rahmat, karunia dan limpahan nikmat-Nya, penulis diberikan kekuatan, kesehatan, dan kemudahan dalam menyelesaikan skripsi yang berjudul “Perbandingan Model Peramalan Kualitas Udara Menggunakan Metode LSTM dan *Hybrid* ARIMA – LSTM (Studi Kasus : Provinsi Sumatera Selatan)”. Dalam proses penelitian ini, peneliti menyadari bahwa terselesaikannya skripsi ini tidak terlepas dari bantuan berbagai pihak. Oleh karena itu, dengan penuh rasa hormat, penulis ingin menyampaikan ucapan terima kasih kepada :

1. Bapak Dr. Eng. Helmy Fitriawan, S.T., M. Sc., selaku Dekan Fakultas Teknik Universitas Lampung;
2. Ibu Herlinawati, S.T., M.T., selaku Ketua Jurusan Teknik Elektro Universitas Lampung
3. Ibu Yessi Mulyani, S.T., M.T., selaku Ketua Program Studi Teknik Informatika Universitas Lampung sekaligus dosen penguji dalam sidang skripsi yang telah memberikan banyak masukan serta saran bagi penulis.
4. Bapak Ir. M. Komarudin, S.T., M.T., selaku Pembimbing Utama yang telah meluangkan waktunya untuk memberikan bimbingan, masukan, semangat dan tambahan ilmu yang berharga selama proses pengerjaan skripsi ini.
5. Ibu Ir. Titin Yulianti, S.T., M. Eng., selaku Pembimbing Pendamping yang telah meluangkan waktunya untuk membimbing, mengarahkan dan memberi masukan serta ilmu yang berharga selama proses pengerjaan skripsi ini.
6. Ibu Resty Annisa, S.ST., M.KOM., selaku dosen pembimbing akademik yang telah memberikan dukungan dan bimbingan selama masa perkuliahan.
7. Seluruh Dosen Program Studi Teknik Informatika Universitas Lampung yang telah memberikan banyak ilmu kepada penulis.
8. Ibu Barikah Utami, PIC Data AQMS Dinas Lingkungan Hidup Prov. Sumatera Selatan, atas bantuan dalam penyediaan data untuk penelitian ini.

9. Mba Rika Asliana, selaku Admin Program Studi Teknik Informatika yang banyak membantu penulis khususnya terkait administrasi.
10. Kedua orang tua dan seluruh keluarga yang telah memberikan doa serta dukungan moril maupun materi
11. Amelia Fitri keluarga sekaligus sahabat kecil yang tumbuh bersama yang selalu mendukung dan menghibur di kala penulis merasa jenuh dan lelah.
12. Ayuk Ayung dan rekannya yang telah membantu dalam kelancaran proses pengerjaan skripsi.
13. Suci, Rapih, Rapli dan Darmisi, sahabat masa kecil yang telah mengukir banyak cerita, memberikan semangat dan dukungan hingga saat ini. Terima kasih sudah bersedia menjadi pendengar setia dalam setiap cerita penulis.
14. Adilah, Agustin dan Azizah teman "IT gurl" sekaligus rekan seperjuangan dibangku perkuliahan, yang selalu memberikan semangat, dukungan dan saling menguatkan dalam setiap langkah yang dilalui bersama.
15. Ridho, Gibran, Muhkito, Afif dan Entin yang telah banyak membantu penulis sejak awal perkuliahan serta Cela Febriyani rekan penelitian yang telah banyak membantu penulis dalam proses pengerjaan skripsi.
16. Alysa Astry rekan seperjuangan yang telah mendengarkan keluh kesah dan memberikan masukan positif kepada penulis dan Rosalena Brigita sahabat magang di DPR RI yang memberikan dukungan hingga saat ini
17. Serta pihak yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari sepenuhnya bahwa penelitian ini masih ada kekurangan karena keterbatasan pengetahuan dan keahlian. Akhir kata, penulis berharap semoga penelitian ini dapat bermanfaat bagi kita semua, Aamiin

Bandar Lampung, 10 Juli 2025

Peneliti,

Nabila Firzariani

DAFTAR ISI

	Halaman
ABSTRAK	i
PERSEMBAHAN	viii
SANWACANA	ix
DAFTAR ISI	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xvi
I. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	4
1.5 Batasan Masalah.....	4
1.6 Sistematika Penulisan Skripsi	5
II. TINJAUAN PUSTAKA	7
2.1 Landasan Teori	7
2.1.1 Indeks Kualitas Udara	7
2.1.2 Pencemaran Udara	11
2.1.3 Peramalan (<i>Forecasting</i>)	11
2.1.4 <i>Data Cleaning</i>	12
2.1.5 Normalisasi Data	13
2.1.6 Uji Korelasi	13
2.1.7 <i>Artificial Intelligence</i>	14
2.1.8 <i>Machine Learning</i>	15
2.1.9 <i>Deep Learning</i>	16
2.1.10 Jaringan Saraf Tiruan	16
2.1.11 <i>Recurrent Neural Network (RNN)</i>	17

2.1.12	<i>Long Short – Term Memory (LSTM)</i>	18
2.1.13	<i>AutoRegressive Integrated Moving Average (ARIMA)</i>	21
2.1.14	<i>Hybrid ARIMA – LSTM</i>	23
2.1.15	Evaluasi Model	24
2.1.16	<i>OSEMN Framework</i>	25
2.2.	Penelitian Terkait	28
III.	METODOLOGI PENELITIAN	33
3.1	Waktu dan Tempat	33
3.2	Alat dan Bahan	34
3.3	Tahapan Penelitian	36
3.3.1	Obtain	37
3.3.2	Scrub	37
3.3.3	Explore	39
3.3.4	Model	40
3.3.5	Interpreter	43
IV.	HASIL DAN PEMBAHASAN	44
4.1	Obtain	44
4.2	Scrub	44
4.3	Explore	55
4.4	Modeling	65
4.5	Interpreter	94
4.6	Analisis dan Pembahasan	99
V.	PENUTUP	102
5.1	Kesimpulan	102
5.2	Saran	103
	DAFTAR PUSTAKA	104
	LAMPIRAN	108

DAFTAR GAMBAR

Gambar 2. 1 Jaringan Saraf Tiruan	16
Gambar 2. 2 Struktur Jaringan Recurrent Neural Network.....	17
Gambar 2. 3 Struktur Jaringan Long-Short Term Memory.....	19
Gambar 2. 4 OSEM Framework	25
Gambar 3. 1 Tahapan Penelitian	36
Gambar 3. 2 Tahapan Obtain.....	37
Gambar 3. 3 Tahapan Scrub	37
Gambar 3. 4 Hasil Penggabungan Data	38
Gambar 3. 5 Tahapan Explore.....	39
Gambar 3. 6 Alur Pembuatan Model LSTM.....	40
Gambar 3. 7 Flowchart Pemilihan Jumlah Lapisan Model LSTM	41
Gambar 3. 8 Flowchart Pemilihan Jumlah Epoch Model LSTM.....	41
Gambar 3. 9 Alur Pembuatan Model Hybrid ARIMA-LSTM	42
Gambar 3. 10 Tahapan Interpreter.....	43
Gambar 4. 1 Hasil Penggabungan Data	45
Gambar 4. 2 Informasi Dataset	46
Gambar 4. 3 Jumlah Missing Value Pada Setiap Variabel	47
Gambar 4. 4 Persentase Missing Value pada dataset	47
Gambar 4. 5 Source Code Mengisi Missing Value Menggunakan Nilai Mean	48
Gambar 4. 6 Source Code Mengisi Missing Value Menggunakan Nilai Median .	49
Gambar 4. 7 Boxplot setiap polutan.....	50
Gambar 4. 8 Source Code Menghitung zscore	50
Gambar 4. 9 Hasil Perhitungan z-score pada dataset dengan imputasi nilai mean	51
Gambar 4. 10 Source Code Fungsi Deteksi Outlier.....	51
Gambar 4. 11 Hasil Perhitungan z – score pada dataset dengan imputasi nilai median	52
Gambar 4. 12 Boxplot Setelah Outlier Diatasi.....	53

Gambar 4. 13 Tabel dataset setelah pre-processing	54
Gambar 4. 14 Hasil Uji Korelasi pada dataset dengan imputasi nilai mean	57
Gambar 4. 15 Hasil Uji Korelasi pada dataset dengan imputasi nilai median	58
Gambar 4. 16 Grafik Polutan Udara PM2.5	59
Gambar 4. 17 Grafik Polutan Udara PM10	60
Gambar 4. 18 Grafik Polutan Udara CO	61
Gambar 4. 19 Grafik Polutan Udara NO2	61
Gambar 4. 20 Grafik Polutan Udara SO2	62
Gambar 4. 21 Grafik Polutan Udara O3	63
Gambar 4. 22 Grafik Indeks Standar Pencemar Udara	63
Gambar 4. 23 Source Code Sliding Window	67
Gambar 4. 24 Ringkasan Arsitektur Model Satu Lapisan LSTM (100 Neuron) ..	68
Gambar 4. 25 Ringkasan Arsitektur Model Satu Lapisan LSTM (256 Neuron) ..	69
Gambar 4. 26 Ringkasan Arsitektur Model Dua Lapisan LSTM (128 dan 64 Neuron)	69
Gambar 4. 27 Ringkasan Arsitektur Model Dua Lapisan LSTM (256 dan 128 Neuron)	70
Gambar 4. 28 Ringkasan Arsitektur Model Tiga Lapisan LSTM	71
Gambar 4. 29 Grafik Loss Satu Lapisan LSTM (100 Neuron)	72
Gambar 4. 30 Grafik Loss satu lapisan LSTM (256 neuron)	73
Gambar 4. 31 Grafik Loss Dua Lapisan LSTM (128 dan 64 Neuron)	74
Gambar 4. 32 Grafik Loss Model Dua Lapisan LSTM (256 dan 128 Neuron)	75
Gambar 4. 33 Grafik Loss Model dengan Tiga Lapisan LSTM	76
Gambar 4. 34 Grafik ACF dan PACF ISPU_MAX	83
Gambar 4. 35 Hasil Model ARIMA pada dataset dengan imputasi nilai mean	84
Gambar 4. 36 Hasil Model ARIMA pada dataset dengan imputasi nilai median .	86
Gambar 4. 37 Grafik nilai residual model ARIMA	88
Gambar 4. 38 Rangkuman Arsitektur Model LSTM Residual	90
Gambar 4. 39 Grafik loss pada residual dengan dataset yang diimputasi nilai mean	91
Gambar 4. 40 Grafik loss pada residual dengan dataset yang diimputasi nilai median	91

Gambar 4. 41 Source Code Rumus Penggabungan Model	92
Gambar 4. 42 Grafik Perbandingan Hasil Prediksi Model LSTM dan Data Aktual	94
Gambar 4. 43 Grafik perbandingan hasil prediksi model ARIMA dan data aktual	95
Gambar 4. 44 Grafik Perbandingan Hasil Prediksi Model Hybrid dan Nilai Aktual	96
Gambar 4. 45 Perbandingan data aktual dan hasil prediksi	97
Gambar 4. 46 Grafik Kinerja Model terhadap Persentase Missing Value	100

DAFTAR TABEL

Tabel 2. 1 kategori rentang nilai ISPU	9
Tabel 2. 2 Konversi nilai konsentrasi parameter ISPU	10
Tabel 2. 3 Kriteria Nilai MAPE	24
Tabel 2. 4 Penelitian Terkait.....	30
Tabel 3. 1 Jadwal kegiatan penelitian	33
Tabel 3. 2 Perangkat keras	34
Tabel 3. 3 Perangkat lunak	34
Tabel 4. 1 Nilai Mean Polutan Udara.....	48
Tabel 4. 2 Nilai Median Polutan Udara.....	49
Tabel 4. 3 Jumlah outlier pada dataset dengan imputasi nilai mean	52
Tabel 4. 4 Jumlah outlier pada dataset dengan imputasi nilai median	53
Tabel 4. 5 Perhitungan Statistik Deskriptif pada dataset dengan imputasi nilai mean	55
Tabel 4. 6 Perhitungan Statistik Deskriptif pada dataset dengan imputasi nilai median.....	56
Tabel 4. 7 Pembagian data	65
Tabel 4. 8 perbandingan hasil nilai sebelum dan sesudah normalisasi data.....	66
Tabel 4. 9 Hasil Evaluasi Model dengan Satu Lapisan LSTM (100 Neuron).....	73
Tabel 4. 10 Hasil Evaluasi Model dengan Satu Lapisan LSTM (256 neuron)	74
Tabel 4. 11 Hasil Evaluasi Model dengan Dua Lapisan LSTM (128 dan 64 Neuron)	75
Tabel 4. 12 Hasil Evaluasi Model dengan Dua Lapisan LSTM (256 dan 128 Neuron)	75
Tabel 4. 13 Hasil Evaluasi Model dengan Tiga Lapisan LSTM	76
Tabel 4. 14 Perbandingan Hasil Evaluasi Setiap Arsitektur Model	77
Tabel 4. 15 Perbandingan Hasil Evaluasi Setiap Jumlah Epoch.....	78
Tabel 4. 16 Evaluasi Metrik Model LSTM	79

Tabel 4. 17 Hasil uji stasioner pada dataset dengan imputasi nilai mean	81
Tabel 4. 18 Hasil uji stasioner pada dataset dengan imputasi nilai median	81
Tabel 4. 19 Perbandingan matriks evaluasi arsitektur model ARIMA.....	84
Tabel 4. 20 Hasil Evaluasi Metrik Model ARIMA	87
Tabel 4. 21 Pembagian data training dan data testing nilai residual	88
Tabel 4. 22 Nilai residual sebelum dan sesudah normalisasi data	89
Tabel 4. 23 Evaluasi metrik model hybrid ARIMA - LSTM	93
Tabel 4. 24 Perbandingan evaluasi metrik	98

I. PENDAHULUAN

1.1 Latar Belakang

Kualitas udara merupakan kondisi udara yang di suatu wilayah dan memiliki dampak bagi kesehatan manusia serta makhluk hidup lainnya. Kualitas udara yang baik akan memberikan dampak positif terhadap kualitas kehidupan manusia dan juga makhluk hidup lainnya begitupun sebaliknya, kualitas udara yang buruk dapat berdampak negatif bagi kehidupan. Menurut *World Health Organization* (WHO) polusi udara adalah bentuk kontaminasi lingkungan baik di dalam maupun di luar ruangan. Terdapat lebih dari 91% penduduk dunia tinggal di daerah dengan kualitas udara yang buruk, dimana tingkat polusi udara pada daerah tersebut melebihi standar dari *World Health Organization* (WHO) [1]. Tingkat polusi udara dapat diukur dengan menggunakan *Air Quality Index* (AQI), AQI adalah standar ukuran yang digunakan secara global untuk menilai kualitas udara serta tingkat polusi udara pada suatu wilayah atau daerah yang dihitung berdasarkan konsentrasi enam polutan antara lain *Particulate Matter 2.5*(PM_{2.5}), *Particulate Matter 10*(PM₁₀), *Karbon Monoksida* (CO), *Sulfur Dioksida* (SO₂), *Nitrogen Dioksida* (NO₂) dan *Ozon*(O₃). Semakin tinggi nilai AQI, maka kualitas udara di wilayah tersebut semakin berbahaya bagi Kesehatan [2].

Di Indonesia, kualitas udara dinilai menggunakan parameter pencemaran udara yang diukur berdasarkan nilai baku mutu udara ambien nasional. Dalam memudahkan penilaian kualitas udara, nilai – nilai parameter pencemaran udara dikonversikan menjadi Indeks Kualitas Udara (IKU) atau biasa disebut juga Indeks Standar Pencemaran Udara (ISPU). ISPU merupakan standar ukuran kualitas udara yang mirip dengan *Air Quality Index* (AQI), hanya saja nilai standar pada ISPU disesuaikan dengan keadaan atau kondisi Indonesia. ISPU digunakan untuk memberikan informasi kepada masyarakat Indonesia mengenai kondisi udara dan dampaknya terhadap kesehatan.

Kota Palembang merupakan ibukota Provinsi Sumatera Selatan yang berfungsi sebagai pusat administrasi dan pendidikan di wilayah tersebut. Sebagai ibukota Provinsi Sumatera Selatan, Kota Palembang mengalami perkembangan yang cukup pesat sehingga terjadi peningkatan aktivitas sehari – hari yang menimbulkan berbagai tantangan yang harus dihadapi salah satunya adalah kondisi lingkungan khususnya kualitas udara. Penurunan kualitas udara Kota Palembang disebabkan oleh berbagai aktivitas masyarakat serta kebakaran hutan. Setiap musim kemarau, Kota Palembang sering kali dilanda kabut asap yang disebabkan oleh kebakaran lahan gambut disekitar wilayahnya. Selain itu, penggunaan kendaraan dan juga aktivitas industri juga memberikan kontribusi terkait penurunan kualitas udara. Berdasarkan data indeks standar pencemaran udara (ISPU) kualitas udara di Provinsi Sumatera Selatan khususnya Kota Palembang berada pada rentang nilai indeks 101 hingga 199 yang artinya kondisi kualitas udara di wilayah tersebut tidak sehat [3]. Kualitas udara buruk dapat menimbulkan berbagai penyakit kesehatan seperti asma, infeksi saluran pernapasan akut (ISPA), jantung hingga kematian. Berdasarkan data dari Badan Pusat Statistika Provinsi Sumatera Selatan, jumlah kasus ISPA meningkat tinggi dari 32.336 kasus pada tahun 2021 menjadi 539.327 kasus pada 2023 [4].

Dalam mengatasi permasalahan tersebut, perlu dilakukan pemeliharaan kualitas udara untuk menjaga kesehatan, kesejahteraan serta melindungi manusia dan makhluk hidup dari dampak negatif udara yang tercemar. Upaya yang dapat dilakukan untuk mengatasi permasalahan tersebut salah satunya adalah melakukan peramalan kualitas udara. Dengan adanya peramalan kualitas udara memungkinkan masyarakat untuk memprediksi kondisi udara di masa depan, sehingga masyarakat dapat mengambil langkah – langkah pencegahan sejak awal. Selain itu, dengan adanya informasi mengenai ramalan kualitas udara masyarakat dapat lebih sadar menjaga kesehatan dan kelestarian lingkungan. Terdapat beberapa penelitian yang telah dilakukan untuk memprediksi kualitas udara di suatu wilayah dengan memanfaatkan algoritma *machine learning* yaitu *forecasting* atau peramalan. Penerapan tersebut digunakan untuk meramalkan kualitas udara dengan memanfaatkan data historis berupa data *time series* yang berisi parameter yang mempengaruhi polusi udara.

Pada penelitian ini model *forecasting* atau peramalan akan dibangun menggunakan dua algoritma yang berbeda. Model peramalan yang pertama akan dibangun dengan menggunakan algoritma Long Short – Term Memory (LSTM) yang merupakan metode *Recurrent Neural Networks* (RNN) yang dirancang untuk mengatasi pola *non – linier* dan kompleksitas yang terdapat pada data kemudian akan dibandingkan dengan algoritma *hybrid ARIMA – LSTM*. Model peramalan dibangun dengan menggunakan dataset yang diperoleh dari Dinas Lingkungan Hidup Provinsi Sumatera Selatan yang berisi nilai konsentrasi polutan pencemaran udara. Penelitian ini diharapkan dapat membantu Pemerintah Provinsi Sumatera Selatan untuk mengambil tindakan lebih cepat dan tepat dalam mengelola serta mengurangi dampak negatif pencemaran udara.

1.2 Rumusan Masalah

Berdasarkan latar belakang, rumusan masalah yang mendasari penelitian adalah sebagai berikut :

1. Bagaimana membangun model peramalan kualitas udara di Provinsi Sumatera Selatan dengan menggunakan metode LSTM dan *hybrid ARIMA – LSTM*?
2. Diantara metode LSTM dan *hybrid ARIMA – LSTM* model manakah yang lebih optimal yang dapat digunakan untuk melakukan peralaman kualitas udara di Provinsi Sumatera Selatan?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut :

1. Melakukan pembuatan model peramalan serta menganalisis pola dan tren kualitas udara di Provinsi Sumatera Selatan dengan menggunakan metode LSTM dan *hybrid ARIMA – LSTM*.
2. Menganalisis dan membandingkan kinerja model LSTM dan *hybrid ARIMA – LSTM* untuk menentukan model mana yang paling optimal dalam melakukan peralaman kualitas udara di Provinsi Sumatera Selatan.

1.4 Manfaat Penelitian

Manfaat dari dilakukannya penelitian ini adalah sebagai berikut :

1. Manfaat Teoritis

Penelitian ini diharapkan dapat menambah wawasan serta menjadi masukan dalam melakukan penelitian mengenai penggunaan metode *hybrid* pada *deep learning* dengan algoritma *AutoRegressive Integrated Moving Average* (ARIMA) dan *Long Short – Term Memory* (LSTM) untuk model peramalan kualitas udara

2. Manfaat Praktis

a. Bagi Peneliti atau akademisi

Penelitian ini diharapkan dapat menjadi referensi bagi para peneliti dalam memahami penerapan metode *hybrid* pada *deep learning* dengan algoritma ARIMA dan LSTM untuk peramalan kualitas udara dan hasil penelitiannya dapat dijadikan sebagai acuan dalam melakukan penelitian serupa.

b. Bagi Masyarakat

Hasil penelitian ini diharapkan dapat memberikan informasi dan menjadi referensi yang bermanfaat bagi masyarakat terkait kualitas udara di Provinsi Sumatera Selatan. Dengan begitu, masyarakat dapat mengambil tindakan pencegahan lebih cepat sehingga dapat terhindar dari dampak negatif polusi udara.

1.5 Batasan Masalah

Batasan masalah dalam penelitian ini mencakup hal – hal berikut :

1. Penelitian berfokus pada analisis dan perbandingan kinerja model LSTM dan *hybrid* ARIMA – LSTM dengan menggunakan nilai MAPE dan RMSE untuk mengetahui performa dari kedua metode.
2. Model dibangun berdasarkan dataset yang didapatkan dari Dinas Lingkungan Hidup dan Pertanahan Provinsi Sumatera Selatan dari tanggal 01 Januari 2020 hingga 12 Desember 2024.

1.6 Sistematika Penulisan Skripsi

Adapun sistematika penulisan skripsi / tugas akhir pada laporan penelitian ini adalah sebagai berikut :

BAB I : PENDAHULUAN

Bab ini berisi mengenai pembahasan latar belakang dan rumusan masalah dilakukannya perbandingan metode LSTM dan *hybrid* ARIMA – LSTM, tujuan dilakukannya perbandingan metode LSTM dan *hybrid* ARIMA – LSTM, membahas batasan masalah dalam perbandingan model peramalan kualitas udara dan sistematika penulisan laporan penelitian.

BAB II : TINJAUAN PUSTAKA

Bab ini berisi dasar – dasar teori yang menjadi landasan dalam penelitian ini serta karya ilmiah yang berkaitan dengan penelitian. Dasar teori yang digunakan pada penelitian ini antara lain indeks kualitas udara, pencemaran udara, peramalan, *artificial intelligence*, *machine learning*, *deep learning* serta algoritma yang digunakan yaitu LSTM dan ARIMA, evaluasi model dan OSEMN *framework*.

BAB III : METODOLOGI PENELITIAN

Bab ini berisi mengenai waktu dan tempat penelitian, alat dan bahan yang diperlukan untuk proses penelitian serta tahapan pembuatan model menggunakan metode OSEMN *framework* yang terdiri dari tahap *obtain* data, *scrub* data, *explore* data, *modeling* dan *interpreter*. Setiap tahapan menjelaskan proses yang dilakukan secara detail.

BAB IV : PEMBAHASAN

Bab ini berisi pembahasan mengenai hasil dari penelitian terkait perbandingan metode LSTM dan *hybrid* ARIMA - LSTM pada peramalan kualitas udara. Hasil yang dipaparkan berupa proses pembuatan serta pengujian model dan nilai error dari setiap metode yang digunakan.

BAB V : KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari hasil penelitian perbandingan metode LSTM dan *hybrid* ARIMA – LSTM dalam meramalkan kualitas udara Kota Palembang, serta saran dari hasil penelitian untuk pengembangan pada penelitian berikutnya.

II. TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 Indeks Kualitas Udara

Indeks kualitas udara merupakan alat ukur yang digunakan untuk menilai atau mengindikasikan tingkat polusi udara atau pencemaran kualitas udara di suatu wilayah. Di Indonesia, Indeks Kualitas Udara dikenal juga dengan Indeks standar pencemaran udara (ISPU) merupakan standar yang digunakan oleh lembaga pemerintah Indonesia untuk mengukur tingkat polusi udara di suatu wilayah. Hasil pengukuran tersebut digunakan untuk memberikan informasi kepada masyarakat mengenai kualitas udara dan dampak yang diterima akibat polusi udara tersebut [5]. ISPU dihitung berdasarkan konsentrasi dari beberapa parameter polutan utama yang ada di udara, antara lain sebagai berikut :

a. *Particulate Matter 2.5 (PM_{2.5})*

PM_{2.5} merupakan partikel halus di udara dengan ukuran diameter kurang dari 2.5 mikrometer, yang dihasilkan dari berbagai sumber yaitu pembakaran bahan bakar fosil, asap kendaraan, dan aktivitas industri. Dengan ukurannya yang sangat kecil memungkinkan partikel tersebut untuk masuk lebih dalam ke paru – paru dan memiliki potensi masuk ke dalam aliran darah yang dapat menyebabkan berbagai penyakit pernapasan hingga kematian dini.

b. *Particulate Matter 10 (PM₁₀)*

PM₁₀ merupakan jenis polutan yang sama dengan PM_{2.5} hanya saja ukurannya lebih besar yaitu sekitar 10 mikrometer. Partikel ini berasal dari berbagai sumber seperti asap kendaraan, debu jalanan, serta aktivitas industri. Karena ukurannya yang tergolong kecil, PM₁₀ dapat masuk ke saluran pernapasan bagian atas dan apabila terhirup dapat menyebabkan iritasi dan dapat memperburuk kondisi kesehatan bagi penderita penyakit asma dan paru – paru kronis [6].

c. Karbon Monoksida (CO)

Karbon Monoksida merupakan senyawa kimia yang tidak memiliki warna dan juga tidak memiliki bau namun sangat berbahaya bagi kesehatan apabila dihirup dalam jumlah yang banyak. Karbon monoksida merupakan senyawa kimia dari pembakaran yang tidak sempurna pada bahan bakar yang dihasilkan oleh kendaraan, akibatnya dapat menyebabkan berbagai gangguan kesehatan seperti terhambatnya aliran oksigen di dalam darah [7].

d. Sulfur Dioksida (SO₂)

Sulfur Dioksida (SO₂) merupakan parameter polutan yang menyebabkan penurunan kualitas udara dengan jumlah paling banyak dibanding dengan parameter yang lain. Sulfur dioksida merupakan polutan yang tidak memiliki warna namun memiliki bau yang sangat tajam yang dihasilkan dari asap kendaraan bermotor dan pembakaran bahan bakar fosil. Ketika sulfur dioksida bereaksi dengan uap air maka senyawa tersebut menjadi H₂SO₄ atau asam sulfat yang merupakan penyebab terjadinya hujan asam [7]. SO₂ dapat menyebabkan sesak napas hingga kematian pada manusia dan juga hewan, sedangkan pada tumbuhan SO₂ menghambat proses fotosintesis.

e. Nitrogen Dioksida (NO₂)

Nitrogen Dioksida merupakan sumber utama dari aerosol nitrat yang membentuk particulate matter dengan ukuran kecil yaitu PM_{2.5}. Nitrogen Dioksida merupakan polutan yang dihasilkan dari kegiatan manusia seperti asap kendaraan, pembakaran sampah, industri dan lain – lain. Polutan ini memiliki warna coklat kemerahan dan juga memiliki bau yang sangat tajam yang dapat merusak paru – paru. Dampak negatif NO₂ terhadap lingkungan adalah merusak tumbuh – tumbuhan [6].

f. Ozon (O₃)

Ozon adalah polutan yang terdapat di permukaan tanah dan merupakan polutan yang cukup berbahaya. Polutan ini merupakan unsur utama penyusun kabut asap yang terbentuk dari reaksi antara sinar matahari dengan nitrogen oksida (NO_x)

dan *volatile organic compounds* (VOC). Ozon dengan konsentrasi yang tinggi dapat menyebabkan berbagai penyakit seperti gangguan pernapasan, serangan jantung, dan kematian [6].

Indeks Standar Pencemar Udara (ISPU) memiliki skala nilai dari 0 sampai dengan 500, nilai yang tinggi menunjukkan kualitas udara yang buruk. Dengan kata lain semakin tinggi nilai ISPU maka semakin buruk kualitas udara di wilayah tersebut. ISPU dibagi menjadi beberapa kategori dari "baik" hingga "berbahaya" dengan resiko yang berbeda setiap tingkatnya[5]. Berikut merupakan rentang nilai ISPU yang dapat dilihat pada tabel 1.

Tabel 2. 1 kategori rentang nilai ISPU

Rentang Angka	Kategori	Status Warna	Keterangan
0 – 50	Baik	Hijau	Tingkat mutu udara yang baik, tidak memberikan efek yang buruk terhadap manusia, hewan ataupun tumbuhan
51 – 100	Sedang	Biru	Tingkat mutu udara yang masih dapat diterima pada kesehatan manusia, hewan ataupun tumbuhan
101 – 200	Tidak Sehat	Kuning	Tingkat mutu udara yang bersifat merugikan manusia, hewan ataupun tumbuhan
201 – 300	Sangat Tidak Sehat	Merah	Tingkat mutu udara yang dapat meningkatkan resiko kesehatan pada sejumlah populasi yang terpapar
≥301	Berbahaya	Hitam	Tingkat mutu udara yang dapat merugikan kesehatan serius pada populasi dan perlu dilakukan penanganan cepat

Menghitung nilai ISPU merupakan aspek penting agar dapat menentukan kategori untuk kualitas udara di suatu wilayah. Nilai ISPU pada penelitian ini dihitung menggunakan dasar perhitungan yang sudah ditetapkan oleh Peraturan

Menteri Lingkungan Hidup dan Kehutanan Republik Indonesia Nomor P.12/MENLHK/SETJEN/KUM.1/7/2020 tentang ISPU. Berikut merupakan rumus untuk menghitung nilai ISPU [5] :

$$I = \frac{I_a - I_b}{X_a - X_b} (X_x - X_b) + I_b \quad (1)$$

Keterangan :

$$I = \text{Nilai ISPU}$$

Ia, Ib = Batas nilai atas & bawah ISPU

Xa, Xb = Batas atas konsentrasi polutan udara

X_x = Hasil pengukuran konsentrasi polutan udara ($\mu\text{g}/\text{m}^3$)

Nilai ambang batas atas dan batas bawah untuk Indeks Standar Pencemar Udara (ISPU) serta konsentrasi masing – masing parameter polutan udara (ambien) telah ditetapkan oleh Kementerian Lingkungan Hidup dan Kehutanan (KLHK) melalui peraturan yang telah berlaku. Penetapan ini bertujuan untuk mengonversi hasil konsentrasi polutan udara menjadi nilai ISPU yang terstandarisasi. Nilai – nilai tersebut digunakan sebagai dasar dalam perhitungan ISPU dan disajikan dalam bentuk tabel konversi sebagai berikut :

Tabel 2. 2 Konversi nilai konsentrasi parameter ISPU

ISPU	24 Jam					
	Partikulat (PM10)	Partikulat (PM2.5)	Sulfur Dioksida (SO2)	Karbon monoksida (CO)	Nitrogen Dioksida (NO2)	Ozon (O3)
0 – 50	50	15,5	52	4000	80	120
51 – 100	150	55,4	180	8000	200	235
101 – 200	350	150,4	400	15000	1130	400
201 – 300	420	250,4	800	30000	2260	800
> 300	500	500	1200	45000	3000	1000

Hasil perhitungan ISPU parameter partikulat (PM10), sulfur dioksida (SO2), karbon monoksida (CO), ozon (O3) dan nitrogen dioksida (NO2), diambil nilai ISPU parameter tertinggi

2.1.2 Pencemaran Udara

Udara menjadi kebutuhan paling utama bagi keberlangsungan makhluk hidup di bumi yang terbentuk dari oksigen dan juga nitrogen. Namun, kualitas udara yang bersih semakin menurun setiap harinya akibat pencemaran udara. Pencemaran udara merupakan keadaan dimana udara telah terkontaminasi oleh berbagai partikel atau zat berbahaya yang dapat memberikan dampak negatif bagi lingkungan, kesehatan manusia, hewan dan makhluk hidup lainnya. Pencemaran udara juga dapat diartikan turunnya kualitas udara yang menyebabkan terjadinya penurunan mutu pada udara sehingga tidak dapat digunakan lagi sebagaimana mestinya. Terjadinya pencemaran udara disebabkan oleh partikel atau zat berbahaya yang dihasilkan dari berbagai kegiatan manusia seperti penggunaan kendaraan bermotor, kebakaran hutan, kegiatan industri dan berbagai kegiatan lainnya sehingga kualitas udara mengalami penurunan. Polusi udara memiliki pengaruh yang besar bagi kesehatan manusia dan makhluk hidup lainnya karena dapat menyebabkan berbagai gangguan kesehatan misalnya radang paru – paru, infeksi saluran pernapasan atas (ISPA) dan gangguan kesehatan lainnya [8].

2.1.3 Peramalan (*Forecasting*)

Peramalan atau *forecasting* merupakan teknik yang digunakan dalam analisa perhitungan dengan melakukan pendekatan kualitatif atau kuantitatif untuk memprediksi kejadian atau peristiwa di masa yang akan datang berdasarkan referensi data masa lalu [6]. Menurut Nasution dan Prasetyawan, peramalan merupakan sebuah proses untuk memprediksi beberapa kebutuhan di masa yang akan datang meliputi kebutuhan dalam ukuran kuantitas, kualitas, waktu serta lokasi untuk memenuhi permintaan barang ataupun jasa. Tujuan dilakukan peramalan untuk mengetahui peristiwa yang akan terjadi di masa yang akan datang sehingga dapat mengambil tindakan yang lebih cepat dan tepat. Selain itu, peramalan dilakukan untuk membantu dalam membuat perencanaan baik jangka pendek maupun jangka panjang.

2.1.4 Data Cleaning

Data *cleaning* atau pembersihan data merupakan proses mengubah atau memperbaiki data mentah menjadi data yang terstruktur dan siap digunakan dalam proses analisis maupun pembuatan model machine learning. Tahap ini bertujuan untuk mengatasi permasalahan nilai yang hilang (*missing value*), mengatasi data yang tidak konsisten (*noise*), mengidentifikasi dan mengatasi data duplikat serta mengidentifikasi dan mengatasi outlier[9].

a. Missing Value

Missing value atau nilai hilang merupakan sebuah kondisi pada sebagian data yang tidak tersedia atau tidak tercatat dalam suatu dataset. Kondisi tersebut dapat disebabkan oleh berbagai faktor, antara lain keterbatasan alat ukur, kegagalan sistem dalam memperoleh data, atau kesalahan dalam tahap pencatatan data. *Missing value* yang tidak diatasi dengan baik dapat mempengaruhi kualitas data, menurun akurasi ataupun performa model peralaman atau prediktif. Terdapat dua pendekatan utama dalam mengatasi *missing value* yaitu penghapusan data dan imputasi. Penghapusan data dapat digunakan apabila *missing value* berjumlah sangat kecil. Sementara itu, apabila *missing value* dalam jumlah yang besar dapat digunakan metode imputasi yaitu mengisi *missing value* menggunakan nilai rata – rata (*mean*) atau nilai tengah (*median*). [10]

b. outlier

Outlier atau pencilan merupakan sebuah data yang nilainya jauh diluar pola distribusi data lainnya[9]. Outlier pada dataset dapat mempengaruhi hasil analisis ataupun kinerja model *machine learning*. Deteksi outlier merupakan bagian penting untuk memastikan kualitas data tetap terjaga. Salah satu metode yang dapat digunakan dalam mengidentifikasi outlier adalah metode *z – score* yaitu mengukur seberapa jauh suatu nilai menyimpang dari rata – rata distribusi data dalam satuan standar deviasi[11]. Nilai *z – score* dihitung dengan menggunakan rumus berikut :

$$Z = \frac{X - \mu}{\sigma} \quad (2)$$

Ket :

Z = nilai *z – score*

X = nilai data

μ = rata – rata data

σ = standar deviasi dari data

Berdasarkan rumus tersebut, semakin besar nilai Z yang dihasilkan maka semakin jauh data dari pusat distribusi. Sebuah data dapat dianggap sebagai outlier apabila nilai Z pada suatu data lebih besar dari 3 atau lebih kecil dari -3.

2.1.5 Normalisasi Data

Normalisasi data merupakan suatu proses transformasi nilai – nilai numerik ke dalam skala tertentu untuk menyamakan rentang nilai antar variabel dan bertujuan untuk mengurangi bias yang disebabkan oleh perbedaan skala antar variabel. Salah satu metode yang biasa digunakan dalam normalisasi data adalah *MinMax Normalization*. *Min-max normalization* merupakan metode yang digunakan untuk mengubah skala pada suatu variabel ke dalam rentang $[0, 1]$ atau $[-1, 1]$. Proses normalisasi menggunakan metode MinMax dapat dilakukan dengan rumus berikut:

$$v' = \frac{v - \min(x)}{\max(x) - \min(x)} \quad (3)$$

Ket :

v = nilai asli x

$\min(x)$ = nilai minimum dari x

v' = nilai hasil normalisasi

$\max(x)$ = nilai maksimum dari x

2.1.6 Uji Korelasi

Uji korelasi merupakan sebuah metode analisis yang biasa digunakan untuk mengetahui dan menganalisis hubungan antara dua variabel. Salah satu metode yang paling umum digunakan untuk mengukur hubungan *linier* antara dua variabel adalah uji korelasi pearson. Nilai koefisien pada uji korelasi pearson berada dalam rentang 1 hingga -1, dimana nilai 1 menunjukkan hubungan positif yang kuat antar variabel sedangkan nilai -1 menunjukkan hubungan negatif yang kuat antar variabel. Uji korelasi pearson dihitung menggunakan rumus berikut :

$$\rho(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4)$$

Ket :

$\rho(X, Y)$ = koefisien korelasi pearson

x_i, y_i = nilai ke – i variabel x dan y

$\bar{x}, \bar{y}$ = nilai rata – rata variabel x dan y

n = jumlah data

2.1.7 Artificial Intelligence

Artificial Intelligence atau dalam bahasa Indonesia dikenal sebagai kecerdasan buatan merupakan sebuah program komputer yang dibuat atau dirancang untuk memproses data, mempelajari data, membuat keputusan serta menyelesaikan masalah yang kompleks. Dari sisi keilmuan kecerdasan buatan (AI) adalah sebuah cabang ilmu komputer yang mempelajari bagaimana menciptakan sebuah sistem cerdas yang berfokus pada pengembangan algoritma dan teknologi serta memungkinkan komputer berperilaku layaknya manusia. Selain itu juga, kecerdasan buatan bertujuan untuk menciptakan sistem yang mampu melakukan tugas-tugas manusia baik di bidang kesehatan, pemrosesan informasi, dan lain sebagainya. Di Indonesia, Kecerdasan Buatan (AI) banyak dimanfaatkan di berbagai bidang antara lain bidang pendidikan, jasa, produk, manufaktur hingga kesehatan [12]. Secara garis besar kecerdasan buatan (AI) dibedakan menjadi 4 kategori yaitu *Thinking Humanly*, *Acting Humanly*, *Thinking Rationally*, *Acting Rationally*. Terdapat beberapa teknik yang dapat dilakukan dalam memecahkan masalah dalam kecerdasan buatan, antara lain :

1. *Searching*, sebuah teknik yang digunakan untuk mencari sebuah solusi dari suatu permasalahan. Contoh penerapan teknik ini pada aplikasi *Google Maps*.
2. *Reasoning*, sebuah teknik yang mengambil kesimpulan berdasarkan fakta. Contoh penerapan teknik yaitu sistem yang digunakan untuk mendiagnosis penyakit.
3. *Planning*, sebuah teknik yang memecahkan masalah dengan membagi masalah menjadi lebih sederhana dan diselesaikan satu per satu.
4. *Learning*, sebuah teknik untuk mempelajari atau meniru cara kerja otak manusia yaitu dengan melakukan proses berulang-ulangan untuk memperoleh pengetahuan.

Kecerdasan Buatan (AI) memiliki hubungan yang sangat erat dengan sistem cerdas, yang mana AI memiliki peran yang cukup penting yaitu sebagai komponen utama yang digunakan untuk menciptakan sistem dengan kemampuan yang dianggap cerdas. Sistem cerdas dirancang untuk memberikan solusi cerdas terhadap

permasalahan yang ada di dunia nyata. Sistem yang dibangun dengan menggunakan kecerdasan buatan (AI) akan menghasilkan sebuah output berupa solusi dari permasalahan berdasarkan pengetahuan yang sudah dipelajari oleh sistem yang dibangun. Salah satu contoh penerapan sistem cerdas adalah *smart home system* yang memanfaatkan AI untuk keamanan dimana aktivitas yang mencurigakan dapat diidentifikasi lebih cepat.

2.1.8 *Machine Learning*

Machine Learning merupakan salah satu cabang dari *Artificial Intelligence* atau kecerdasan buatan yang berfokus pada pengembangan sistem berdasarkan data yang dipelajarinya. *Machine Learning* menggabungkan beberapa cabang ilmu dalam proses pengembangannya diantaranya adalah ilmu komputer, ilmu matematika dan ilmu statistik. *Machine learning* adalah sebuah aplikasi komputer yang memanfaatkan algoritma matematika untuk menghasilkan sebuah prediksi melalui proses pembelajaran. Terdapat dua penerapan utama dalam *machine learning* yaitu klasifikasi dan prediksi. Klasifikasi merupakan sebuah metode yang digunakan untuk mengkategorikan atau mengklasifikasikan sebuah objek berdasarkan pada ciri tertentu sedangkan prediksi merupakan sebuah metode yang digunakan untuk memprediksi hasil dari sebuah inputan berdasarkan data yang sudah dipelajari sebelumnya. *Machine Learning* terbagi menjadi beberapa algoritma yang berbeda antara lain sebagai berikut :

1. *Supervised Learning*, algoritma yang mengambil kesimpulan berdasarkan data pelatihan yang sudah diberikan label.
2. *Unsupervised Learning*, Algoritma yang bertujuan untuk meningkatkan kinerja dengan melibatkan data yang tidak diketahui dan model yang dihasilkan untuk menemukan pola datanya.
3. *Semi-supervised Learning*, Algoritma yang digunakan untuk melatih model pada data yang berjumlah besar tanpa banyak label.
4. *Reinforcement Learning*, algoritma yang mempelajari bagaimana mengambil keputusan terbaik dalam suatu situasi.

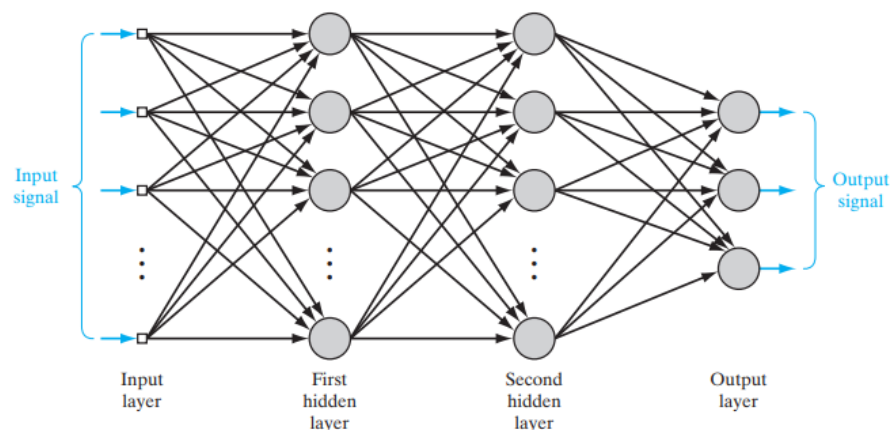
Machine Learning menghasilkan model yang menyimpan pengetahuannya ke dalam bentuk parameter. Kelebihan memanfaatkan machine learning adalah kemampuan yang dimiliki dalam mempelajari serta melakukan pembaharuan terhadap pengetahuan yang dimiliki secara otomatis.

2.1.9 Deep Learning

Deep Learning merupakan cabang dari *machine learning* yang algoritmanya mirip dengan cara kerja struktur otak manusia sehingga *deep learning* mampu untuk mengidentifikasi pola yang rumit yang ada di dalam data seperti teks, foto, suara dan lain sebagainya. Teknologi ini sangat cocok untuk digunakan untuk mengolah data mentah dan juga menciptakan pola yang dapat digunakan membantu proses pengambilan keputusan. Dalam prosesnya, *deep learning* memanfaatkan beberapa lapisan yang terhubung secara hierarkis sehingga data yang diterima akan diproses dan disederhanakan terlebih dahulu sebelum diteruskan ke lapisan berikutnya[13].

2.1.10 Jaringan Saraf Tiruan

Jaringan Saraf Tiruan merupakan sebuah metode pemrosesan yang mirip dengan sel syaraf yang mirip dengan cara kerja otak manusia. Jaringan Saraf Tiruan atau *Artificial Neural Network* terdiri dari beberapa unit pemrosesan yang saling terhubung untuk menyimpan pengetahuan yang diperoleh dari pengalaman sehingga dapat digunakan ketika hal tersebut dibutuhkan.



Gambar 2. 1 Jaringan Saraf Tiruan[14]

Pengetahuan yang diperoleh pada jaringan saraf tiruan merupakan hasil dari proses pembelajaran. Setiap neuron yang terhubung memiliki bobot yang dikenal sebagai bobot sinaptik fungsinya adalah menyimpan pengetahuan yang sudah di pelajari sebelumnya. Terdapat 3 lapisan penyusun jaringan saraf tiruan yaitu lapisan input, lapisan tersembunyi (hidden layer) dan lapisan output [14]. Secara mematis, proses kerja sebuah neuron pada JST adalah sebagai berikut :

$$y = f \left(\sum_{i=1}^n w_i x_i + b \right)$$

Ket :

y = output

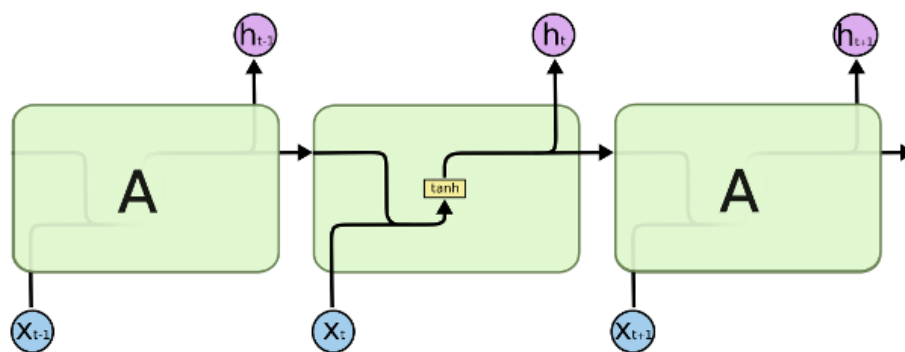
x_i = input ke – i

b = bias

w_i = bobot ke – i

2.1.11 Recurrent Neural Network (RNN)

Recurrent Neural Network atau jaringan saraf berulang merupakan pengembangan dari *Artificial Neural Network* yang dapat digunakan pada data sekuensial dan melakukan pemrosesan secara berulang. RNN terdiri dari neuron yang saling terhubung dan membentuk sebuah siklus pada jaringannya sehingga outputnya dapat digunakan sebagai input pada langkah selanjutnya Informasi masa lalu yang tersimpan pada memori memberikan pengaruh untuk informasi yang baru[15].



Gambar 2. 2 Struktur Jaringan Recurrent Neural Network

Berdasarkan strukturnya, RNN memiliki *hidden state* (h_t) yang berfungsi sebagai memori untuk menyimpan hasil dari proses perhitungan antara nilai input (x_t) dan langkah sebelumnya (h_{t-1}). Secara matematis, proses komputasi pada algoritma RNN adalah sebagai berikut :

$$h_t = \tanh(W_x x_t + W_h h_{t-1} + b)$$

$$y_t = \tanh(W_o h_t + c)$$

Ket :

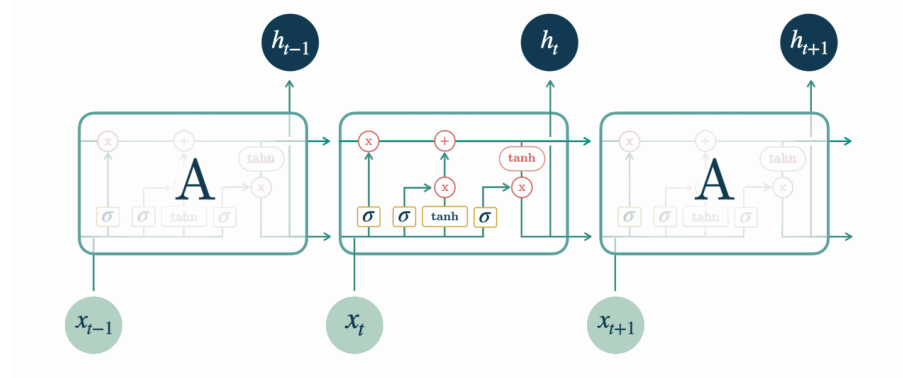
x_t	= input ke - t	h_t	= <i>hidden state</i> ke - t
y_t	= output ke - t	$W_x W_h W_o$	= bobot input, hidden, output
b, c	= bias	tanh	= fungsi aktivasi

Secara teori, RNN memiliki arsitektur yang dapat menyelesaikan permasalahan ketergantungan jangka panjang, namun pada penerapannya RNN mengalami kesulitan dalam mempertahankan informasi pada langkah sebelumnya dengan baik karena permasalahan *vanishing gradient*.

2.1.12 Long Short – Term Memory (LSTM)

Long – Short Term Memory (LSTM) merupakan turunan dari algoritma *Recurrent Neural Network* (RNN) yang digunakan untuk mengatasi keterbatasan RNN dalam memprediksi data yang membutuhkan penyimpanan informasi dalam jangka panjang. Algoritma LSTM dilatih menggunakan *backpropagation through time* sehingga algoritma ini cukup efektif untuk mengatasi masalah “*vanishing gradient*” atau “*exploding gradient*” yang terjadi dalam algoritma RNN. Pada algoritma LSTM data atau informasi dapat disimpan dalam jangka waktu yang lama dan akan dilakukan pembuangan pada beberapa data atau informasi yang sudah tidak relevan. Pada Algoritma LSTM setiap neuron terdapat beberapa *gates* atau gerbang yang akan mengatur memori untuk setiap *neuronnya*. *Gates* atau gerbang merupakan mekanisme yang digunakan oleh algoritma LSTM untuk menyeleksi informasi atau data yang masuk ke dalam *cell* yang berfungsi untuk mengatur informasi yang masuk dari waktu ke waktu[13].

Arsitektur pada algoritma LSTM antara lain *memory cell*, *cell state*, *input gate*, *output gate*, dan *forget gate*. Tiga *gates* yang terdapat pada algoritma LSTM juga berfungsi untuk melindungi dan mengontrol *cell state*.



Gambar 2. 3 Struktur Jaringan *Long-Short Term Memory*

a. *Memory Cell*

Memory cell memiliki fungsi yang sama dengan neuron yang dipelajari pada Jaringan Saraf Tiruan. *Memory cell* merupakan komponen pada *Algoritma LSTM* yang bertugas untuk menyimpan dan memperbarui informasi serta mempertahankan informasi yang masih relevan [16].

b. *Forget Gate*

Forget gate merupakan gerbang pertama yang dioperasikan pada model *Long – Short Term Memory* (LSTM), pada gerbang ini informasi yang penting akan dipertahankan dan informasi yang dianggap sudah tidak penting atau tidak relevan akan dibuang sehingga data yang tersedia tetap *up – to – date* sesuai dengan kebutuhan[13]. Pada gerbang atau *gate* ini terdapat dua input yang diterima yaitu h_{t-1} dan x_t . *Output* yang akan dihasilkan pada *forget gate* adalah angka dengan rentang 0 sampai 1, apabila *output* berupa *sigmoid* mendekati angka 0 maka informasi tersebut akan dilupakan, sebaliknya apabila *output* mendekati angka 1 maka informasi tersebut akan dipertahankan [17]. Fungsi yang digunakan untuk menghasilkan nilai 0 hingga 1 dalam *forget gate* diuraikan pada persamaan berikut :

$$f_t = \sigma (W_f \cdot [h_{t-1}, x_t] + b_f) \quad (5)$$

Ket :

f_t	= <i>Forget gate</i>	h_{t-1}	= <i>hidden state cell</i> sebelum
W_f	= bobot <i>forget gate</i>	x_t	= nilai <i>input</i>
σ	= fungsi aktivasi <i>sigmoid</i>	b_f	= bias <i>forget gate</i>

c. *Input Gate*

Setelah nilai *forget gate* diperoleh, maka langkah selanjutnya adalah *input gate*. *Input gate* memiliki dua bagian yaitu “*layer sigmoid*” dan “*tanh layer*”. *Layer sigmoid* digunakan untuk memperbarui informasi, sedangkan *tanh layer* digunakan untuk menyimpan nilai atau informasi baru ke dalam *memory cell*[16]. Fungsi tersebut diuraikan pada persamaan berikut :

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (6)$$

$$\check{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (7)$$

Ket :

i_t	= <i>input gate</i>	$\tanh$	= fungsi aktivasi <i>tanh</i>
W_i	= bobot <i>input gate</i>	W_c	= bobot <i>candidate gate</i>
b_i	= bias <i>input gate</i>	b_c	= bias <i>candidate gate</i>
C_t	= <i>candidate gate</i>		

d. *Cell State*

Cell state memiliki fungsi untuk mengganti nilai yang ada pada *memory cell* sebelumnya dengan nilai *memory cell* yang baru[16]. Operasi pada *cell state* dapat dilihat pada uraian persamaan berikut :

$$c_t = f_t * c_{t-1} + i_t * \check{C}_t \quad (8)$$

Ket :

c_t	= <i>cell state</i>	f_t	= <i>forget gate</i>
i_t	= <i>input gate</i>	c_{t-1}	= nilai <i>cell state</i> sebelum
$\check{C}_t$	= <i>candidate gate</i>		

e. *Output Gate*

Output gate memiliki fungsi sebagai gerbang terakhir yang akan menghasilkan informasi sebagai *output* dari model LSTM. *Output gate* bertugas untuk menentukan nilai *memory cell* yang akan dikeluarkan dengan menggunakan fungsi aktivasi *sigmoid*, selanjutnya nilai *cell state* diolah menggunakan fungsi aktivasi *tanh* yang memungkinkan informasi yang dihasilkan lebih stabil[16]. Rumus yang digunakan untuk menghitung nilai *output gate* diuraikan pada persamaan berikut :

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (9)$$

$$h_t = o_t * \tanh(c_t) \quad (10)$$

Ket :

o_t = *output gate*

h_t = *hidden state*

W_o = bobot *output gate*

b_o = bias *output gate*

c_t = *cell gate*

2.1.13 *AutoRegressive Integrated Moving Average (ARIMA)*

AutoRegressive Integrated Moving Average (ARIMA) merupakan metode statistik yang biasa digunakan untuk membuat model peramalan dengan pendekatan deret waktu. Algoritma ini dikembangkan oleh *George Box* dan *Gwilym Jenkins* dan sering disebut dengan ARIMA Box – Jenkins. Algoritma ini memanfaatkan data historis dan data sekarang serta pola tren pada variabel dependen untuk menghasilkan peramalan jangka pendek yang akurat, namun algoritma ini tidak tepat untuk peramalan jangka panjang. Model ARIMA merupakan kombinasi dari model *Autoregressive (AR)* dan *Moving Average (MA)*, hal ini dikarenakan kedua model tersebut tidak dapat mengatasi permasalahan tren dan ketidakstasioneran secara bersamaan pada data time series[18].

a. *Autoregressive (AR)*

Model AR merupakan model dengan nilai variabel baru dipengaruhi oleh nilai variabel dari waktu sebelumnya (nilai variabel di masa lalu), yang digunakan sebagai prediktor untuk memprediksi nilai saat ini. Model AR dilambangkan

dengan ordo p yang disingkat menjadi AR(p) atau ARIMA($p,0,0$). Berikut persamaan dari model AR :

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t \quad (11)$$

Ket :

Y_t = deret waktu stasioner

e_t = sisa waktu ke- t

ϕ_p = parameter AR ke- p

b. *Moving Average (MA)*

Model MA menghitung hubungan antara kesalahan prediksi (*error*) sebelumnya dan kesalahan prediksi pada nilai saat ini, sehingga nilai yang diprediksi saat ini dipengaruhi oleh error pada nilai sebelumnya. Model MA dilambangkan dengan ordo q yang disingkat menjadi MA(q) atau ARIMA($0,0,q$). Berikut persamaan dari model MA :

$$Y_t = e_t - \phi_1 e_{t-1} - \phi_2 e_{t-2} - \dots - \phi_q e_{t-q} \quad (12)$$

Ket :

Y_t = deret waktu stasioner

e_t = nilai error

ϕ_q = parameter MA ke- q

Model ARIMA terdiri dari tiga komponen utama yaitu *Autoregressive (AR)*, *Integrated (I)*, dan *Moving Average (MA)*. Untuk membuat model menggunakan algoritma ARIMA data masukan harus stasioner yang artinya rata – rata dan varians dari data harus konstan dari waktu ke waktu. Apabila data tersebut tidak stasioner maka perlu dilakukan penyesuaian data dengan mengoreksi ketidakstasioneran data tersebut agar dapat digunakan untuk membangun model ARIMA. Untuk mengatasi data yang tidak *stasioner* tersebut dilakukanlah sebuah proses *differencing* sehingga model menjadi terintegrasi. Pada model ARIMA terdapat tiga parameter yaitu p , d , dan q yang dinotasikan menjadi ARIMA(p,d,q) dimana parameter p merupakan nilai AR, parameter d sebagai nilai perbedaan, dan parameter q sebagai nilai MA[19]. Parameter tersebut akan disesuaikan dan digunakan untuk mengatasi berbagai tren dan ketidakstasioneran yang terdapat dalam data *time series*. Dengan

begitu dapat diketahui bahwa model ARIMA merupakan model yang bersifat fleksibel sehingga dapat digunakan untuk berbagai jenis data *time series*. Bentuk umum persamaan model ARIMA yaitu sebagai berikut :

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_q Y_{t-q} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q} + e_t$$

Atau

$$\phi_p(B)(1-B)^d Y_t = \theta_q(B) a_t \quad (13)$$

Ket :

$\theta_p(B)$ = *Moving Average* B = *Backshift Operator* ($B y_t = y_{t-1}$)

$\phi_q(B)$ = *Autoregressive* a_t = sisa waktu ke-t

d = *Differencing*

2.1.14 Hybrid ARIMA – LSTM

Hybrid model merupakan sebuah pendekatan yang menggabungkan dua atau lebih algoritma yang berbeda dengan tujuan untuk meningkatkan akurasi pada model yang dibuat serta mengatasi keterbatasan dari masing – masing metode. Sebagaimana yang diketahui bahwa setiap algoritma memiliki kekurangan dan kelebihan masing – masing, dengan menggabungkan keduanya diharapkan metode tersebut dapat saling melengkapi. ARIMA adalah metode statistik yang dirancang untuk melakukan analisis pada data stationer, selain itu ARIMA memiliki komputasi yang ringan dan mudah dipahami. Namun, model ini memiliki keterbatasan dalam menangani pola data yang bersifat *non – linier* dan kompleks. Untuk mengatasi keterbatasan tersebut, model ARIMA biasa digabungkan dengan model *deep learning* seperti *Long Short – Term Memory* (LSTM) yang memiliki kemampuan dalam mengenali pola non – linier. Akan tetapi, model *deep learning* membutuhkan waktu pelatihan yang lebih lama serta komputasi yang lebih besar[20].

b. *Root Mean Square Error (RMSE)*

Root Mean Squared Error (RMSE) adalah sebuah parameter yang biasa digunakan untuk mengevaluasi kinerja model dengan menghitung selisih antara nilai yang diprediksi dengan nilai sebenarnya[21]. Rumus persamaan untuk menghitung nilai RMSE adalah sebagai berikut :

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (x_t - f_t)^2}{n}} \quad (16)$$

Ket :

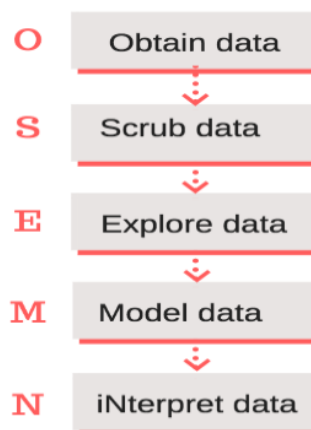
n = jumlah data

f_t = nilai hasil prediksi

x_t = nilai aktual

2.1.16 OSEMN Framework

OSEMN *framework* merupakan tahapan penelitian yang memiliki struktur kerja sistematis dan biasa digunakan untuk pengembangan proyek atau pemecahan masalah pada bidang data science dalam skala besar. OSEMN *framework* memiliki tahapan pengerjaan yang cukup runtut dimulai dari pengumpulan data hingga interpretasi data.



Gambar 2. 4 OSEMN Framework[22]

a. *Obtain*

Obtain Data merupakan tahap awal dalam melakukan penelitian dalam proyek data *science* yaitu mengidentifikasi dan mengumpulkan data yang dibutuhkan sebagai bahan untuk penelitian. Data yang dikumpulkan harus diambil dari sumber yang relevan dan dapat digunakan sesuai dengan tujuan analisis[23]. Pada tahap ini, pengumpulan data dapat memanfaatkan web API atau mengunduh file excel atau csv. yang dapat diperoleh dengan cara tradisional.

b. *Scrub Data*

Setelah mendapatkan data yang dibutuhkan, tahap selanjutnya adalah *scrubbing* data. Scrub data merupakan tahapan untuk mengubah data mentah menjadi data yang siap untuk digunakan dalam analisis data. Pada tahap ini dilakukan pemrosesan data berupa pembersihan data yang tidak relevan dengan penelitian, mengubah format data ke bentuk yang dapat dipahami oleh mesin, mengisi nilai pada data yang kosong, serta menghapus data duplikat. Dengan melakukan data scrubbing, data menjadi siap untuk dianalisis dan memberikan hasil yang lebih akurat[22].

c. *Explore Data*

Explore data merupakan tahapan untuk menemukan pola, mencari informasi serta mengidentifikasi data yang akan digunakan dalam melakukan proses analisis serta pembuatan hipotesis[22]. Tahapan ini bertujuan untuk memahami pola dan karakteristik dari data sebelum dilakukan analisis mendalam. Dalam memahami pola dan karakteristik data dilakukan dengan berbagai cara seperti pembuatan visualisasi data, pengujian statistik deskriptif, uji korelasi, serta analisis eksplorasi lainnya. Pada tahap explore data, proses penghapusan missing value dapat dilakukan untuk mengetahui frame data terdistribusi dengan normal. Oleh karena itu, tahap ini menjadi hal penting yang akan digunakan untuk menentukan metode yang tepat untuk digunakan pada proses analisis.

d. Model

Setelah dilakukan explore data, tahapan selanjutnya adalah modelling atau pemodelan data. Pada machine learning, model merupakan representasi matematis yang dapat digunakan untuk membuat prediksi ataupun melakukan klasifikasi[22]. Tahapan pemodelan data dilakukan dengan memilih algoritma machine learning yang sesuai dengan karakteristik dataset yang dimiliki. Pada tahap ini, digunakan berbagai algoritma pemodelan untuk mendapatkan model yang lebih akurat dalam membuat prediksi atau keputusan berdasarkan data yang tersedia.

e. *Interpreter*

Tahap akhir dari OSEMN adalah interpretasi data. Tahap ini merupakan tahap menginterpretasikan atau menafsirkan model dan data sehingga dapat disajikan dalam bentuk yang mudah dipahami oleh audiens atau masyarakat. Pada tahap ini, dilakukan penarikan kesimpulan serta evaluasi hasil dari proses pemodelan data[22].

2.2. Penelitian Terkait

Dalam pelaksanaan penelitian ini, terdapat berbagai studi literatur yang dijadikan sebagai dasar acuan dalam melakukan penulisan. Pemanfaatan literatur ini bertujuan untuk memperluas pemahaman terkait topik yang sedang dikaji serta mengembangkan aspek – aspek baru berdasarkan penelitian – penelitian sebelumnya.

Sebagai dasar acuan dalam penyusunan skripsi ini, penelitian yang pertama kali menjadi referensi adalah penelitian yang dilakukan oleh Muhammad Enrico Suryo Wicaksono dkk. [6] Pada penelitian ini model dibuat menggunakan *Long Short – Term Memory* dan *Support Vector Regression* untuk melakukan peramalan kualitas udara di Jakarta Pusat.. Model terbaik yang dihasilkan pada penelitian ini adalah model LSTM dengan *windows size* 12 dan menghasilkan rata – rata MAPE sebesar 12,15% dan rata – rata RMSE sebesar 0,0941.

Selanjutnya terdapat penelitian yang dilakukan oleh Hikmah, dkk. [24] yaitu membangun model ARIMA dan RNN untuk peramalan data cuaca ekstrim di Indonesia. Pada penelitian ini ditambahkan metode optimasi Adam dan SGD serta fungsi aktivasi reLu, tanh, sigmoid dan gaussian untuk meningkatkan kemampuan model dalam mempelajari data yang kompleks. Pemodelan menggunakan ARIMA-RNN-ADAM menghasilkan akurasi sebesar 97,61% untuk *data training* sedangkan ARIMA-RNN-SGD akurasi *data training* yang dihasilkan sebesar 95,50%. Pada *data testing* model ARIMA-RNN-ADAM memberikan akurasi 87,76% sedangkan model ARIMA-RNN-SGD memberikan akurasi sebesar 75,90%.

Penelitian berikutnya dilakukan Nida Faoziatun Khusna, dkk. [25] membangun model menggunakan ARIMA untuk peramalan kualitas udara di Semarang. Pada penelitian ini dilakukan uji kestasioneran kemudian hasilnya menyatakan bahwa data tidak *stasioner* lalu dilakukan proses *differencing* setelah itu diuji kembali menggunakan uji *Augmented Dickey-Fuller* (ADF) sehingga dihasilkan *P-value* 0,01. Pemilihan model terbaik dilakukan dengan melihat nilai *Akaike's Information Criterion* (AIC) paling kecil sehingga model yang diperoleh adalah model ARIMA(1,1,1) dengan nilai AIC sebesar 2164.69

Penelitian yang dilakukan oleh Laras Wiranda, dkk.[26] bertujuan untuk memprediksi penjualan produk menggunakan LSTM. Pada penelitian ini model

LSTM dibangun dengan komposisi 90% sebagai data latih dan 10% sebagai data uji dengan range interval $[-1, 1]$. Setelah dilakukan epoch sebanyak 1500 kali, diperoleh nilai RMSE sebesar 13,762 dan nilai MAPE sebesar 12%.

Penelitian yang dilakukan oleh Fariza Tolesh, dkk. [27] terkait peramalan migrasi internasional di Kazakhstan dengan menggunakan metode ARIMA. Model ARIMA yang diperoleh adalah ARIMA(0, 1, 0) dengan nilai RMSE 12190.954 dan MAPE 20.026. Model ini efektif dalam mengatasi data yang *staisoner*, namun kurang efektif untuk digunakan dalam peramalan jangka panjang.

Penelitian terkait peramalan juga dilakukan oleh Emmanuel Dave, dkk. [20] Penelitian ini menggunakan data ekspor bulanan Indonesia yang berfokus pada pengembangan model *hybrid* ARIMA – LSTM. Hasil dari penelitian ini menunjukkan bahwa model *hybrid* ARIMA – LSTM lebih unggul dibandingkan dengan model individual masing – masing model ARIMA dan LSTM, dimana model *hybrid* ARIMA – LSTM memiliki nilai evaluasi kesalahan yang rendah yaitu nilai MAPE sebesar 7.38% dan RMSE sebesar 1.66×10^{13} .

Penelitian selanjutnya dilakukan oleh Hossein Abbasimehr, dkk. [28] yang bertujuan untuk meningkatkan akurasi model peramalan menggunakan metode LSTM dalam meramalkan data penjualan produk. Sebelumnya model dibangun menggunakan beberapa metode yaitu ARIMA, ANN, *Exponential Smoothing*, SVM lalu baru model dibangun menggunakan LSTM. Hasil yang diperoleh dari penelitian ini menyatakan bahwa model LSTM memiliki nilai RMSE dan SMAPE paling rendah diantara model yang lain yaitu sebesar 2961.52 dan 0.1208 yang artinya kesalahan prediksi pada model LSTM relatif rendah.

Selanjutnya, penelitian yang dilakukan oleh Furizal Alfian Ma'arid, dkk.[29] bertujuan untuk membuat model peramalan polutan PM10 pada AQI menggunakan ARIMA dan GRU. Algoritma ARIMA pada penelitian tersebut digunakan untuk mengatasi permasalahan outlier ataupun data yang hilang, lalu model GRU digunakan sebagai model utama. Hasil evaluasi untuk model GRU yang diperoleh yaitu RMSE 3.2170 dan MSE 15.6457, nilai evaluasi yang rendah menunjukkan performa model yang baik dan dapat melakukan peramalan dengan akurat.

Kemudian, terdapat penelitian yang dilakukan oleh Jiahui Duan, dkk.[19] yang bertujuan untuk membangun model prediksi menggunakan ARIMA – CNN –

LSTM serta *Dung Beetle Optimizer* untuk mengatasi optimasi *hyperparameter*. Penelitian ini menggabungkan model ARIMA untuk mengenali pola *linear* dari dataset dan model CNN-LSTM digunakan untuk mengenali pola *non-linear*. Model ARIMA-CNN-LSTM mengalami peningkatan akurasi ketika ditambahkan algoritma *Dung Beetle Optimizer* (DBO) dimana nilai RMSE berkurang 34.53%, MAE berkurang 34% dan R^2 meningkat sebesar 1.64%.

Selain itu, terdapat penelitian yang dilakukan oleh Maryem Rhanoui, dkk.[18] Membahas penggunaan metode ARIMA *random walk* dan LSTM *neural network* dalam peramalan. Hasil evaluasi model ARIMA yaitu RMSE 0.239; MAE 0.139; dan MSE 0.057 dan untuk model LSTM hasil evaluasi yang diperoleh yaitu RMSE 0.222; MAE 0.119; dan MSE 0.049. Berdasarkan hasil evaluasi tersebut model LSTM lebih unggul dibandingkan dengan model ARIMA. Hasil prediksi pada penelitian ini menunjukkan bahwa anggaran akan bertumbuh dengan stabil untuk beberapa waktu ke depan, hasil prediksi tersebut memberikan wawasan kepada pihak yang berwewenang untuk melakukan pengambilan keputusan.

Tabel 2. 4 Penelitian Terkait

No	Peneliti	Metode	Hasil
1.	Muhammad Enrico Suryo Wicaksono, dkk. (2023) [6]	LSTM dan SVR	Metode LSTM menghasilkan rata – rata MAPE sebesar 12,15% dan rata – rata RMSE sebesar 0,0941 dan Metode SVR menghasilkan rata – rata MAPE sebesar 16,19% dan rata – rata RMSE sebesar 0.1666.
2.	Hikmah, dkk. (2023) [24]	ARIMA dan RNN	Model ARIMA menghasilkan nilai RMSE paling tinggi dibandingkan dengan model ARIMA – RNN – ADAM dan ARIMA – RNN – SGD.
3.	Nida Faoziatun Khusna, dkk. (2023) [25]	ARIMA	Model ARIMA(1,1,1) menghasilkan nilai MAPE sebesar 13.714%

No	Peneliti	Metode	Hasil
4.	Laras Wiranda, Mujiono Sadikin. (2019) [26]	LSTM	Model LSTM dibangun dengan 90% data latih dan 10% data uji dengan <i>range interval</i> $[-1, 1]$. Setelah dilakukan epoch sebanyak 1500 diperoleh hasil nilai RMSE sebesar 13,762 dan nilai MAPE sebesar 12%.
5.	Fariza Tolesh, dkk. (2024) [27]	ARIMA	Model ARIMA(0, 1, 0) memperoleh RMSE 12190.954 dan MAPE 20.026. Model ini efektif dalam mengatasi dataset yang bersifat stasioner.
6.	Emmanuel Dave, dkk. (2021)[20]	Hybrid Model ARIMA – LSTM	<i>Hybrid</i> Model ARIMA – LSTM menghasilkan akurasi yang lebih baik dibandingkan dengan masing masing model individual ARIMA dan LSTM. <i>Hybrid</i> Model ARIMA – LSTM memiliki kinerja yang baik dengan perolehan nilai MAPE sebesar 7.38% dan RMSE 1.66×10^{13}
7.	Hossein Abbasimehr, dkk. (2020) [28]	LSTM, ANN, SVM, KNN, RNN	Model di evaluasi menggunakan RMSE dan SMAPE. Nilai RMSE dan SMAPE paling rendah terdapat pada model LSTM yaitu sebesar 2961.52 dan 0.1208 untuk data pengujian dibandingkan dengan model lain
8.	Furizal, Alfian Ma'arid, dkk. (2024) [29]	ARIMA dan GRU	Model ARIMA dibangun untuk mengatasi permasalahan outlier atau data kosong. Hasil evaluasi model GRU yaitu RMSE 3.2170 dan MSE 15.6457.
9.	Jiahui Duan, dkk. (2023) [19]	ARIMA, CNN dan LSTM	Model ARIMA-CNN-LSTM mengalami peningkatan akurasi ketika ditambahkan algoritma <i>Dung Beetle Optimizer</i> (DBO) dimana nilai RMSE berkurang 34.53%, MAE

No	Peneliti	Metode	Hasil
			berkurang 34% dan R^2 meningkat sebesar 1.64%
10.	Maryem Rhanoui, dkk. (2019) [18]	ARIMA random walk dan LSTM neural network	Hasil evaluasi model ARIMA yaitu RMSE 0.239; MAE 0.139; dan MSE 0.057 dan untuk model LSTM hasil evaluasi yang diperoleh yaitu RMSE 0.222; MAE 0.119; dan MSE 0.049

Pada Tabel 2. 4 terdapat beberapa penelitian terkait peralaman dengan menggunakan metode ARIMA, LSTM, GRU, RNN dan lain – lain. Penelitian tersebut menunjukkan metode LSTM memiliki akurasi yang lebih unggul dibandingkan dengan metode lain. Selain itu, penggabungan model ARIMA – LSTM pada studi kasus data ekspor bulanan indonesia menunjukkan peningkatan akurasi dibandingkan dengan model individual LSTM. Berdasarkan penelitian – penelitian yang sudah dilakukan tersebut, belum pernah dilakukan penelitian terkait peramalan kualitas udara untuk studi kasus kualitas udara Provinsi Sumatera Selatan dengan menggunakan metode LSTM ataupun metode *hybrid* ARIMA – LSTM. Penelitian yang akan dilakukan yaitu peramalan kualitas udara Provinsi Sumatera Selatan dengan menggunakan model LSTM dan membandingkannya dengan metode *hybrid* ARIMA – LSTM. Perbandingan ini bertujuan untuk mengetahui metode yang memiliki performa terbaik dalam meramalkan kualitas udara di Provinsi Sumatera Selatan.

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat

Penelitian dan Pembuatan skripsi ini dilakukan pada :

Waktu : September 2024 – Maret 2025

Tempat : Laboratorium Teknik Digital Universitas Lampung

Berikut ini adalah jadwal kegiatan penelitian yang akan dilakukan :

Tabel 3. 1 Jadwal kegiatan penelitian

[illegible]

3.2 Alat dan Bahan

3.2.1 Alat

Alat yang dibutuhkan dalam merancang model peramalan kualitas udara adalah sebagai berikut :

a. Perangkat Keras

Perangkat keras yang dibutuhkan dalam proses penelitian ini dapat dilihat pada tabel berikut :

Tabel 3. 2 Perangkat keras

No.	Nama Perangkat	Spesifikasi	Kegunaan
1	Laptop ASUS X415EP	Intel(R) Core(TM) i7 – 1165G7 CPU @2.80GHz 2.80 GHz	Digunakan sebagai media untuk merancang algoritma peramalan atau <i>forecasting</i> dan juga pembuatan laporan penelitian

b. Perangkat Lunak

Perangkat lunak yang dibutuhkan dalam perancangan algoritma forecasting dapat dilihat pada tabel berikut :

Tabel 3. 3 Perangkat lunak

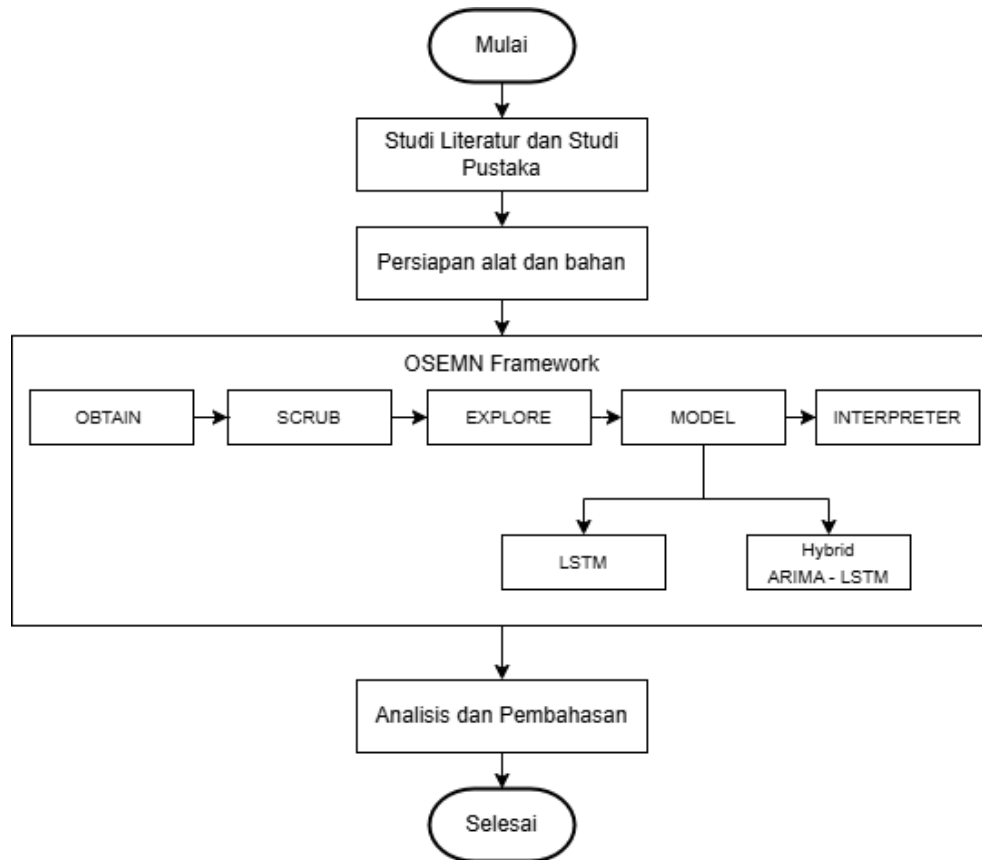
No	Nama Perangkat Lunak	Spesifikasi	Kegunaan
1	Python	3.12.0	Digunakan dalam proses pembuatan model LSTM dan ARIMA – LSTM.
	a. Pandas	1.4.2	Untuk melakukan analisis data termasuk pengolahan data dalam bentuk tabel.
	b. Numpy	1.23.5	Untuk melakukan operasi array multidimensi dan juga komputasi numerik

No	Nama Perangkat Lunak	Spesifikasi	Kegunaan
	c. Statsmodels	0.12.2	Untuk melakukan uji hipotesis, eksplorasi data dan membuat model statistik.
	d. Scikit-learn	1.3.0	Untuk membuat <i>machine learning</i> dan menyediakan berbagai algoritma.
	e. Matplotlib	3.5.2	untuk membuat visualisasi data seperti <i>histogram</i> , <i>scatter</i> , dll. untuk analisis data.
	f. Seaborn	0.12.2	memiliki fungsi sama seperti matplotlib dengan versi yang lebih mudah untuk digunakan.
	g. Keras	2.11.0	untuk jaringan saraf tiruan sehingga memungkinkan digunakan untuk membuat model <i>deep learning</i> .
	h. Tensorflow	2.9.1	untuk membuat <i>machine learning</i> dan <i>deep learning</i> yang dikembangkan oleh google.
2.	Visual Studio Code	1.82.2	<i>Editor</i> kode yang mendukung berbagai bahasa pemrograman
	a. Jupyter Notebook	6.5.4	untuk menulis dan menjalankan kode menggunakan bahasa pemrograman python.
3.	Microsoft Excel	2019	untuk melakukan pengolahan data dan penggabungan data

3.2.2 Bahan

Bahan yang digunakan pada penelitian ini merupakan data kualitas udara Provinsi Sumatera Selatan yang diperoleh dari Dinas Lingkungan Hidup dan Pertanian Provinsi Sumatera Selatan. Dataset terdiri dari tujuh kolom yang mencakup berbagai variabel antara lain tanggal dan waktu, CO, NO₂, O₃, PM_{2.5}, PM₁₀, dan SO₂ dari tahun Januari 2020 hingga Desember 2024.

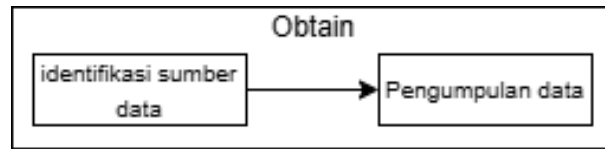
3.3 Tahapan Penelitian



Gambar 3. 1 Tahapan Penelitian

Penelitian ini dilakukan dengan beberapa tahapan yang dapat dilihat pada Gambar 3. 1 tahapan tersebut dimulai dengan melakukan studi literatur yang bertujuan untuk mempelajari dan memahami teori dasar serta penelitian terkait yang telah dilakukan sebelumnya sesuai dengan permasalahan yang akan diteliti. Tahapan selanjutnya yaitu mempersiapkan alat dan bahan yang akan digunakan pada saat melakukan penelitian. Setelah itu, penelitian dilanjutkan menggunakan metode pengembangan yang biasa digunakan dalam data *science* yaitu OSEM^N Framework, metode ini digunakan untuk proses pengolahan data. OSEM^N framework merupakan tahapan atau proses yang terdiri dari lima tahapan yaitu *Obtain*, *Scrub*, *Explore*, *Model*, dan *Interpreter*[22]. Setiap tahapan memiliki peran penting untuk memastikan proses pengolahan data dilakukan secara menyeluruh sehingga memberikan hasil yang *valid*. Setelah setiap tahapan pada OSEM^N Framework selesai dilakukan, tahapan dilanjutkan dengan melakukan penulisan laporan serta menganalisis hasil penelitian yang telah dilakukan.

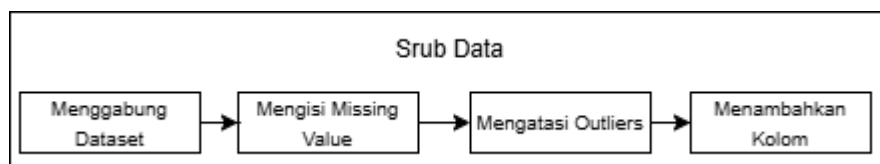
3.3.1 Obtain



Gambar 3. 2 Tahapan Obtain

Pada tahapan ini dilakukan pengumpulan data udara ambien Provinsi Sumatera Selatan. Pengumpulan data dilakukan dengan mengajukan permohonan permintaan data ke Dinas Lingkungan Hidup dan Pertanahan Provinsi, data diambil dari website AQMS yang hanya dapat di akses oleh PIC Stasiun Pemantauan Kualitas Udara. Dataset yang diperoleh dalam format excel sebanyak 30 file yang berisi laporan konsentrasi masing – masing polutan udara, dimana setiap file berisi data yang terdiri dari kolom waktu dan hari untuk setiap bulannya selama 1 tahun dan nilai konsentrasi polutannya. Contoh struktur data pada file tersebut dapat dilihat pada lampiran 2. Selanjutnya, data tersebut akan diolah sebelum digunakan sebagai data *input* sistem prediksi kualitas udara.

3.3.2 Scrub



Gambar 3. 3 Tahapan Scrub

Data yang telah berhasil dikumpulkan, selanjutnya dilakukan *scrub* data sebagai tahap pemrosesan data (*preprocessing*). Pada tahap ini dilakukan penggabungan data dengan menggunakan microsoft excel pada data setiap polutan ke dalam satu file. Data tersebut digabung menjadi sebuah dataset yang memiliki 7 kolom yaitu *Datetime*, *Particulate Matter 2.5(PM_{2.5})*, *Particulate Matter 10(PM₁₀)*, *Karbon Monoksida (CO)*, *Sulfur Dioksida (SO₂)*, *Nitrogen Dioksida (NO₂)* dan *Ozon(O₃)*. Setelah data berhasil digabungkan, tahapan berikutnya adalah

melakukan pembersihan data dengan mengatasi atau mengisi *missing value* serta mengatasi outliers yang terdapat pada dataset. *Missing value* yang terdapat pada dataset dapat diatasi dengan menghapus atau mengisi nilainya dengan menggunakan nilai mean atau median[10]. Hal yang sama juga dilakukan dalam mengatasi outliers yang dimiliki oleh dataset. Outliers merupakan data yang nilainya jauh berbeda dari nilai lainnya pada suatu kumpulan data, outlier ini dapat berupa nilai yang jauh lebih tinggi ataupun lebih rendah dibanding dengan nilai lainnya yang dapat disebabkan karena kesalahan pengukuran dan faktor lainnya. Setelah *missing value* ataupun outliers berhasil diatasi, proses dilanjutkan dengan menambahkan sebuah kolom, penambahan kolom tersebut bertujuan untuk mengetahui nilai ISPU berdasarkan data yang dimiliki dan juga akan digunakan untuk melakukan peramalan kualitas udara. Proses ini dilakukan menggunakan bahasa pemrograman python dengan tools *jupyter notebook*. Tujuan dilakukan proses ini adalah untuk mengubah data mentah menjadi data yang siap diolah dan meningkatkan kualitas dataset. Hasil penggabungan data menjadi sebuah dataset dapat dilihat pada Gambar 3. 4.

	Datetime	PM2.5	PM10	CO	SO2	NO2	O3
0	2020-01-01 00:00:00.000	22.0	22.0	231.0	NaN	18.0	27.0
1	2020-01-01 00:30:00.000	19.0	19.0	229.0	NaN	18.0	28.0
2	2020-01-01 01:00:00.000	16.0	17.0	230.0	NaN	19.0	29.0
3	2020-01-01 01:30:00.000	9.0	14.0	231.0	NaN	19.0	30.0
4	2020-01-01 02:00:00.000	13.0	14.0	231.0	NaN	19.0	30.0
...	...	...	...	...	...	...	...
86779	2024-12-12 21:30:00.018	NaN	NaN	NaN	NaN	NaN	NaN
86780	2024-12-12 22:00:00.018	NaN	NaN	NaN	NaN	NaN	NaN
86781	2024-12-12 22:30:00.018	NaN	NaN	NaN	NaN	NaN	NaN
86782	2024-12-12 23:00:00.018	NaN	NaN	NaN	NaN	NaN	NaN
86783	2024-12-12 23:30:00.018	NaN	NaN	NaN	NaN	NaN	NaN

Gambar 3. 4 Hasil Penggabungan Data

3.3.3 Explore



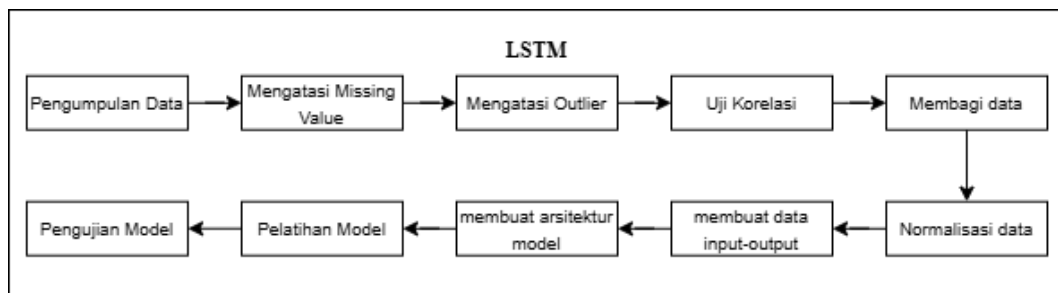
Gambar 3. 5 Tahapan Explore

Tahapan selanjutnya adalah *explore* data, dimana data yang telah diolah sebelumnya akan dilakukan eksplorasi dengan tujuan untuk memahami karakteristik yang dimiliki oleh dataset tersebut. Pada tahap *explore* data terdapat beberapa langkah yang dilakukan antara lain menghitung *statistik deskriptif* yaitu nilai mean, median, nilai maksimum dan minimum dan nilai standar deviasi. Proses selanjutnya yaitu melakukan uji korelasi antar variabel pada dataset dengan memanfaatkan metode matriks korelasi Pearson yang bertujuan untuk menghitung atau mengukur hubungan yang dimiliki antar dua variabel[22]. Hasil uji korelasi ditampilkan dengan menggunakan *heatmap* yang disediakan oleh *library seaborn*. Setelah itu, dilakukan analisis untuk memahami serta mengidentifikasi sebaran pada dataset dengan memanfaatkan visualisasi data. Visualisasi data ditampilkan dengan menggunakan diagram garis untuk mempermudah dalam memahami data. Visualisasi dengan menggunakan diagram garis dilakukan untuk melihat pola perubahan yang terdapat pada nilai konsentrasi polutan dari waktu ke waktu. Dengan melakukan proses eksplorasi diharapkan dapat memperoleh pemahaman yang kuat mengenai dataset yang akan digunakan untuk membuat sebuah model yang akurat.

3.3.4 Model

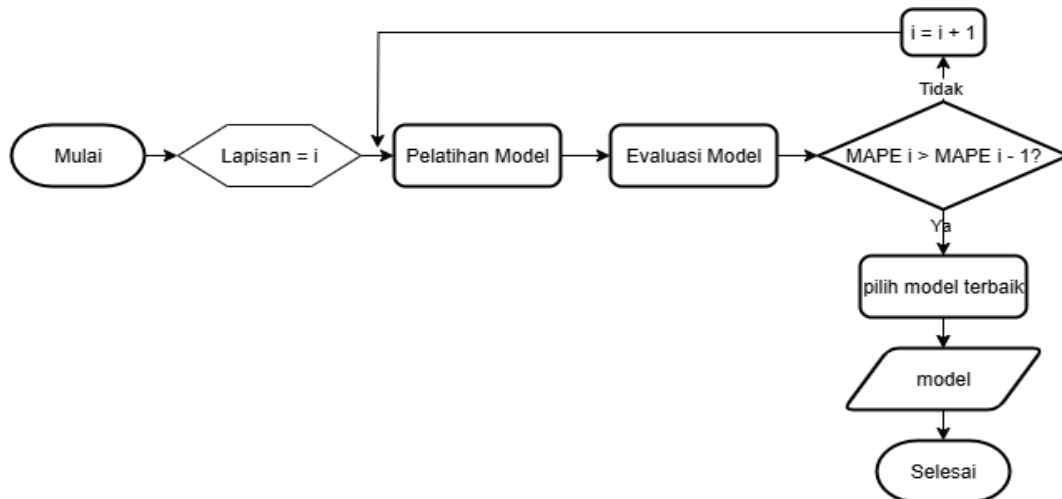
Tahap selanjutnya, pembuat model berdasarkan data yang telah dilakukan proses eksplorasi. Pada tahap pembuatan model akan menggunakan metode *deep learning* yaitu *Long Short-Term Memory* (LSTM) dan *Hybrid ARIMA – LSTM*. Kedua algoritma tersebut diimplementasikan dengan menggunakan bahasa pemrograman *python*.

a. *Long Short-Term Memory* (LSTM)



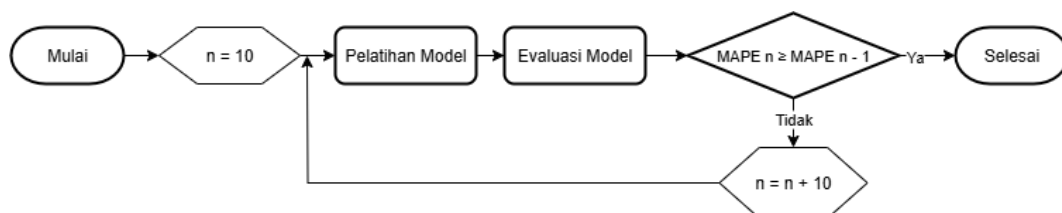
Gambar 3. 6 Alur Pembuatan Model LSTM

Pada proses pembuatan model LSTM, tahap pertama adalah pengumpulan data serta dilakukan *preprocessing* data yaitu mengatasi *missing value*, mengatasi outlier dan melakukan uji korelasi. Setelah dilakukan tahap eksplorasi data selanjutnya, dilakukan proses pembagian dataset ke dalam tiga bagian yaitu data latih (80%) untuk proses pelatihan model, data validasi (10%) untuk mengevaluasi kinerja model selama proses pelatihan dan data uji (10%) untuk mengevaluasi kinerja akhir model. Proses pembuatan model LSTM yang dimulai dengan normalisasi data kemudian dilakukan pembagian data *input – output*. Selanjutnya, dilakukan perancangan atau pembuatan arsitektur model LSTM dengan menentukan jumlah lapisan dan jumlah *epoch* yang optimal. Proses penentuan lapisan model dan jumlah *epoch* dapat dilihat pada flowchart Gambar 3. 7 dan Gambar 3. 8. Tahap selanjutnya dilakukan pelatihan model secara menyeluruh menggunakan data latih. Tahap ini bertujuan agar model mengenali pola dalam data. Setelah itu, model dievaluasi dengan menggunakan evaluasi metrik yaitu *Root Mean Squared Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE).



Gambar 3. 7 Flowchart Pemilihan Jumlah Lapisan Model LSTM

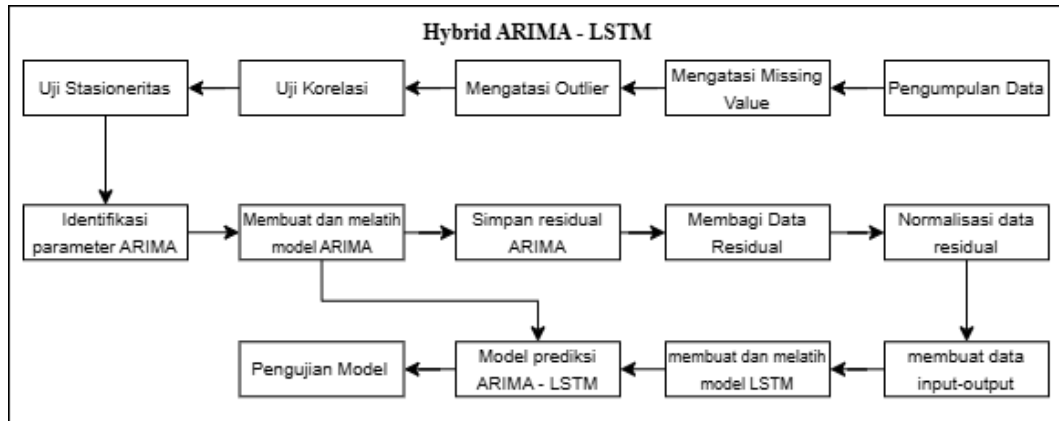
Gambar 3. 7 merupakan flowchart yang menunjukkan proses penentuan jumlah lapisan model untuk arsitektur model LSTM. Proses dimulai dengan menentukan lapisan awal yang diikuti dengan pelatihan evaluasi model. Jika nilai MAPE ke- i lebih kecil dari nilai MAPE sebelumnya ($i - 1$), maka akan dilakukan penambahan lapisan pada model lalu pelatihan dan evaluasi model diulang. Namun, jika nilai MAPE tidak mengalami penurunan ($\text{MAPE } i > \text{MAPE } i - 1$), maka proses dihentikan dan model yang memiliki performa terbaik akan dipilih.



Gambar 3. 8 Flowchart Pemilihan Jumlah Epoch Model LSTM

Gambar 3. 8 merupakan flowchart yang menunjukkan proses penentuan jumlah *epoch* model LSTM. Proses penentuan jumlah *epoch* dimulai dengan inisialisasi *epoch* awal, dilanjutkan dengan pelatihan dan evaluasi model. Jika nilai MAPE ke- n lebih kecil dari nilai MAPE sebelumnya ($n - 1$), maka jumlah *epoch* akan ditambah dan pelatihan diulang. Namun, jika tidak terjadi penurunan nilai MAPE ($\text{MAPE } n > \text{MAPE } n-1$), maka proses dihentikan dan jumlah epoch yang menunjukkan performa optimal yang akan dipilih.

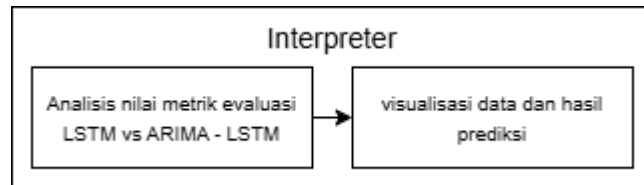
b. *Hybrid* ARIMA – LSTM



Gambar 3. 9 Alur Pembuatan Model *Hybrid* ARIMA-LSTM

Dalam membuat model *hybrid* ARIMA – LSTM, tahapan awal adalah melakukan pengumpulan data dan dilanjutkan dengan melakukan *preprocessing* data yang meliputi penanganan *missing value*, outlier, melakukan uji korelasi serta uji stasioneritas. Tahapan selanjutnya adalah menentukan parameter untuk model ARIMA. Setelah itu, dilakukan pembuatan arsitektur dan pelatihan model ARIMA, model tersebut akan menghasilkan nilai *residual*. Kemudian nilai *residual* tersebut akan disimpan untuk digunakan dalam pembuatan model LSTM. Tahap berikutnya dilakukan normalisasi pada data *residual* dan menentukan data *input – output*, setelah itu dilakukan pembuatan arsitektur model LSTM. Kemudian, model LSTM yang dibangun menggunakan nilai *residual* akan digabungkan dengan model ARIMA untuk membuat model *hybrid* ARIMA – LSTM. Setelah proses penggabungan model ARIMA – LSTM berhasil dibuat, pada tahap akhir akan dilakukan evaluasi model menggunakan RMSE dan MAPE.

3.3.5 Interpreter



Gambar 3. 10 Tahapan Interpreter

Tahap *Interpreter* merupakan tahap akhir dalam metode OSEM *Framework*. Pada tahap ini akan dilakukan analisis terkait hasil dari pengujian model LSTM dan juga model *hybrid* ARIMA – LSTM yang telah dilakukan evaluasi secara mendalam. Evaluasi model dilakukan dengan menghitung metrik evaluasi MAPE dan RMSE untuk mengukur tingkat kesalahan sehingga dapat memahami seberapa baik model dalam melakukan prediksi. Selanjutnya, dilakukan visualisasi hasil prediksi antara model LSTM dan *hybrid* ARIMA – LSTM untuk mempermudah pemahaman terkait performa antara kedua model yang dihasilkan.

V. PENUTUP

5.1 Kesimpulan

Berdasarkan hasil dan analisis dari penelitian terkait perbandingan metode LSTM dan metode *hybrid* ARIMA – LSTM dalam peramalan kualitas udara di Provinsi Sumatera Selatan diperoleh beberapa kesimpulan sebagai berikut :

1. Penelitian ini berhasil membangun dua model peramalan kualitas udara di Provinsi Sumatera Selatan yaitu model LSTM dan model *hybrid* ARIMA – LSTM. Model LSTM berhasil dibangun dengan arsitektur dua lapisan yang terdiri dari 256 dan 128 neuron. Sementara itu, model *hybrid* dibangun dengan menggabungkan model ARIMA(2, 0, 3) dan model LSTM 2 lapisan (256 dan 128 neuron), model ARIMA digunakan untuk menghasilkan nilai residual yang akan dijadikan input pada model LSTM.
2. Berdasarkan hasil evaluasi menggunakan metrik RMSE dan MAPE, model LSTM memiliki performa yang lebih baik dibandingkan model *hybrid* ARIMA – LSTM. Model LSTM menghasilkan nilai RMSE sebesar 5,4746 dan MAPE sebesar 4,04%. Sedangkan model *hybrid* ARIMA – LSTM menghasilkan nilai RMSE sebesar 5,7482 dan MAPE sebesar 4,89%. Selisih nilai RMSE sebesar 0,2736 dan MAPE sebesar 0,85% menunjukkan bahwa model LSTM lebih akurat dan konsisten dalam melakukan peramalan kualitas udara, sehingga model LSTM adalah model yang paling optimal dalam penelitian ini.

5.2 Saran

Adapun saran yang dapat diberikan dari penelitian yang telah dilakukan untuk penelitian selanjutnya adalah sebagai berikut :

1. Pada penelitian selanjutnya disarankan untuk menggunakan teknik optimasi seperti Grid Search atau metode lainnya agar proses pelatihan model tidak memakan waktu yang lama. Selain itu, waktu komputasi perlu diperhatikan sebagai salah satu aspek evaluasi untuk memastikan model tidak hanya akurat namun juga efisien dalam pengimplementasiannya
2. Pada penelitian selanjutnya disarankan untuk mencoba kombinasi model lain dalam membangun model *hybrid* seperti menggabungkan model LSTM dengan model *deep learning* lain agar dapat memperoleh akurasi prediksi yang lebih baik.

DAFTAR PUSTAKA

- [1] D. Hidajat, F. G. Tilana, dan I. G. B. S. A. Kusuma, “Dampak Polusi Udara terhadap Kesehatan Kulit,” *Unram Med. J.*, vol. 12, no. 4, Des 2023, doi: 10.29303/jku.v12i4.1021.
- [2] T. S. Anggraini, H. Irie, A. D. Sakti, dan K. Wikantika, “Machine learning-based global air quality index development using remote sensing and ground-based stations,” *Environ. Adv.*, vol. 15, hlm. 100456, Apr 2024, doi: 10.1016/j.envadv.2023.100456.
- [3] “Berita Bappeda Provinsi Sumsel.” Diakses: 8 Desember 2024. [Daring]. Tersedia pada: <http://bappeda.sumselprov.go.id/berita/173/udara-di-palembang-tidak-sehat-lagi>
- [4] B. P. S. P. S. Selatan, “Jumlah Kasus Penyakit Menurut Jenis Penyakit - Tabel Statistik.” Diakses: 20 Desember 2024. [Daring]. Tersedia pada: <https://sumsel.bps.go.id/id/statistics-table/2/MzY4IzI=/jumlah-kasus-penyakit-menurut-jenis-penyakit.html>
- [5] A. E. Putra dan T. Rismawan, “Klasifikasi Kualitas Udara Berdasarkan Indeks Standar Pencemaran Udara (ISPU) Menggunakan Metode Fuzzy Tsukamoto,” *Coding J. Komput. Dan Apl.*, vol. 11, no. 2, hlm. 190, Sep 2023, doi: 10.26418/coding.v11i2.58704.
- [6] M. E. S. Wicaksono, G. M. Arya Sasmita, dan I. P. A. Eka Pratama, “Peramalan Kualitas Udara di Kota Jakarta Pusat dengan Metode Long Short-Term Memory dan Support-Vector Regression,” *JITTER J. Ilm. Teknol. Dan Komput.*, vol. 4, no. 1, hlm. 1576, Jan 2023, doi: 10.24843/JTRTI.2023.v04.i01.p03.
- [7] S. Siburian, *Pencemaran Udara dan Emisi Gas Rumah Kaca*. Jakarta: Kreasi Cendekia Pustaka, 2020.

- [8] S. Candrasari *dkk.*, “Pemulihan Dampak Pencemaran Udara bagi Kesehatan Masyarakat Indonesia,” *Prof. J. Komun. Dan Adm. Publik*, vol. 10, no. 2, Des 2023, doi: 10.37676/professional.v10i2.5417.
- [9] J. Han, M. Kamber, dan J. Pei, “Data Preprocessing,” dalam *Data Mining*, Elsevier, 2012, hlm. 83–124. doi: 10.1016/B978-0-12-381479-1.00003-4.
- [10] S. García, J. Luengo, dan F. Herrera, *Data Preprocessing in Data Mining*, vol. 72. dalam *Intelligent Systems Reference Library*, vol. 72. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-10247-4.
- [11] V. Aggarwal, V. Gupta, P. Singh, K. Sharma, dan N. Sharma, “Detection of Spatial Outlier by Using Improved Z-Score Test,” dalam *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India: IEEE, Apr 2019, hlm. 788–790. doi: 10.1109/ICOEI.2019.8862582.
- [12] K. R. Ririh, N. Laili, A. Wicaksono, dan S. Tsurayya, “Studi Komparasi Dan Analisis Swot Pada Implementasi Kecerdasan Buatan (Artificial Intelligence) Di Indonesia,” *J. Tek. Ind.*, vol. 15, no. 2, hlm. 122–133, Jun 2020, doi: <https://doi.org/10.14710/jati.15.2.122-133>.
- [13] J. Nurhakiki dan Y. Yahfizham, “Studi Kepustakaan: Pengenalan 4 Algoritma Pada Pembelajaran Deep Learning Beserta Implikasinya,” *J. Pendidik. Berkarakter*, vol. 2, hlm. 279–281, Jan 2024, doi: <https://doi.org/10.51903/pendekar.v2i1.598>.
- [14] S. S. Haykin dan S. S. Haykin, *Neural networks and learning machines*, 3rd ed. New York: Prentice Hall, 2009.
- [15] M. Lukoševičius dan H. Jaeger, “Reservoir computing approaches to recurrent neural network training,” *Comput. Sci. Rev.*, vol. 3, no. 3, hlm. 127–149, Agu 2009, doi: 10.1016/j.cosrev.2009.03.005.
- [16] N. Selle, N. Yudistira, dan C. Dewi, “Perbandingan Prediksi Penggunaan Listrik dengan Menggunakan Metode Long Short Term Memory (LSTM) dan Recurrent Neural Network (RNN),” *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 9, no. 1, hlm. 155–162, Feb 2022, doi: 10.25126/jtiik.2022915585.
- [17] R. Yotenka dan F. F. El Huda, “Implementasi Long Short-Term Memory Pada Harga Saham Perusahaan Perkebunan Di Indonesia,” *Unisda J. Math.*

- Comput. Sci. UJMC*, vol. 6, no. 01, hlm. 9–18, Jun 2020, doi: 10.52166/ujmc.v6i01.1927.
- [18] M. Rhanoui, S. Yousfi, M. Mikram, dan H. Merizak, “Forecasting Financial Budget Time Series: ARIMA Random Walk vs LSTM Neural Network,” *IAES Int. J. Artif. Intell. IJ-AI*, vol. 8, no. 4, hlm. 317, Des 2019, doi: 10.11591/ijai.v8.i4.pp317-327.
- [19] J. Duan, Y. Gong, J. Luo, dan Z. Zhao, “Air-quality prediction based on the ARIMA-CNN-LSTM combination model optimized by dung beetle optimizer,” *Sci. Rep.*, vol. 13, no. 1, Jul 2023, doi: 10.1038/s41598-023-36620-4.
- [20] E. Dave, A. Leonardo, M. Jeanice, dan N. Hanafiah, “Forecasting Indonesia Exports using a Hybrid Model ARIMA-LSTM,” *Procedia Comput. Sci.*, vol. 179, hlm. 480–487, 2021, doi: 10.1016/j.procs.2021.01.031.
- [21] A. Primawati, I. S. Sitanggang, A. Annisa, dan D. A. Astuti, “Perbandingan Kinerja LSTM dan Prophet untuk Prediksi Deret Waktu (Studi Kasus Produksi Susu Sapi Harian),” *J. Edukasi Dan Penelit. Inform. JEPIN*, vol. 9, no. 3, hlm. 428, Des 2023, doi: 10.26418/jp.v9i3.72031.
- [22] K. Dineva dan T. Atanasova, “Osemn Process For Working Over Data Acquired By Iot Devices Mounted In Beehives,” *Curr. Trends Nat. Sci.*, vol. 7, no. 13, hlm. 47–53, 2018.
- [23] F. T. Br Sitepu, V. A. Sirait, dan R. Yunis, “Analisis Runtun Waktu Untuk Memprediksi Jumlah Mahasiswa Baru Dengan Model Prophet Facebook,” *Paradig. - J. Komput. Dan Inform.*, vol. 23, no. 1, Mar 2021, doi: 10.31294/p.v23i1.9756.
- [24] H. Hikmah, A. Asrirawan, A. Apriyanto, dan N. Nilawati, “Peramalan Data Cuaca Ekstrim Indonesia Menggunakan Model ARIMA dan Recurrent Neural Network,” *Jambura J. Math.*, vol. 5, no. 1, Feb 2023, doi: 10.34312/jjom.v5i1.17496.
- [25] N. F. Khusna, S. Aulia, S. Amaria, A. Rahmah, A. Sanmas, dan F. Fauzi, “Peramalan Kualitas Udara di Semarang Menggunakan Metode Autoregressive Integrated Moving Average (ARIMA),” *Pros. Semin. Nas. UNIMUS*, vol. 6, Okt 2023.

- [26] L. Wiranda dan M. Sadikin, “Penerapan Long Short Term Memory Pada Data Time Series Untuk Memprediksi Penjualan Produk Pt. Metiska Farma,” *J. Nas. Pendidik. Tek. Inform. JANAPATI*, vol. 8, hlm. 184–196, 2019, doi: <https://doi.org/10.23887/janapati.v8i3.19139>.
- [27] F. Tolesh dan S. Biloshchytska, “Forecasting international migration in Kazakhstan using ARIMA models,” *Procedia Comput. Sci.*, vol. 231, hlm. 176–183, 2024, doi: 10.1016/j.procs.2023.12.190.
- [28] H. Abbasimehr, M. Shabani, dan M. Yousefi, “An optimized model using LSTM network for demand forecasting,” *Comput. Ind. Eng.*, vol. 143, Mei 2020, doi: 10.1016/j.cie.2020.106435.
- [29] Furizal, A. Ma’arif, I. Suwarno, A. Masitha, L. Aulia, dan A.-N. Sharkawy, “Real-Time Mechanism Based on Deep Learning Approaches for Analyzing the Impact of Future Timestep Forecasts on Actual Air Quality Index of PM10,” *Results Eng.*, vol. 24, Des 2024, doi: 10.1016/j.rineng.2024.103434.
- [30] B. Kim, E. Kim, S. Jung, M. Kim, J. Kim, dan S. Kim, “PM2.5 Concentration Forecasting Using Weighted Bi-LSTM and Random Forest Feature Importance-Based Feature Selection,” *Atmosphere*, vol. 14, no. 6, hlm. 968, Jun 2023, doi: 10.3390/atmos14060968.
- [31] W. Wei dkk., “Application of a Combined Model with Autoregressive Integrated Moving Average (ARIMA) and Generalized Regression Neural Network (GRNN) in Forecasting Hepatitis Incidence in Heng County, China,” *PLOS ONE*, vol. 11, no. 6, hlm. e0156768, Jun 2016, doi: 10.1371/journal.pone.0156768.
- [32] M.-C. Wu, G.-F. Lin, dan H.-Y. Lin, “The effect of data quality on model performance with application to daily evaporation estimation,” *Stoch. Environ. Res. Risk Assess.*, vol. 27, no. 7, hlm. 1661–1671, Okt 2013, doi: 10.1007/s00477-013-0703-4.