

**Penerapan *Data Analytics* untuk Prediksi dan Visualisasi Kualitas Udara
Dalam Ruangan Berbasis LSTM dalam Mendukung Pengambilan
Keputusan**

(Skripsi)

Oleh

MUHAMMAD AFIF RAFI SYAIIM

2115061096



**FAKULTAS TEKNIK
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG**

2025

**Penerapan *Data Analytics* untuk Prediksi dan Visualisasi Kualitas Udara
Dalam Ruangan Berbasis LSTM dalam Mendukung Pengambilan
Keputusan**

**Oleh
MUHAMMAD AFIF RAFI SYAIIM**

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

Penerapan Data Analytics untuk Prediksi dan Visualisasi Kualitas Udara Dalam Ruang Berbasis LSTM dalam Mendukung Pengambilan Keputusan

Oleh

Muhammad Afif Rafi Syaiim

Kualitas udara dalam ruangan berpengaruh besar terhadap kesehatan, kenyamanan, dan produktivitas, terutama pada ruang tertutup seperti kelas dan laboratorium. Permasalahan yang sering muncul adalah sulitnya melakukan pemantauan secara berkelanjutan serta kebutuhan akan sistem prediksi yang dapat membantu pengambilan keputusan preventif. Penelitian ini bertujuan untuk mengembangkan sistem prediksi dan visualisasi kualitas udara dalam ruangan yang informatif dan mudah dipahami. Metodologi yang digunakan adalah *Obtain, Scrub, Explore, Model, dan Interpret* (OSEMN) untuk pengolahan data historis, dengan pemodelan menggunakan algoritma *Long Short-Term Memory* (LSTM) untuk memprediksi tren polutan. Evaluasi model dilakukan menggunakan *Mean Absolute Percentage Error* (MAPE) sebagai metrik utama, dengan hasil 6,90% untuk Sekolah-1 dan 11,16% untuk Sekolah-2. Selain itu, *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE) juga menunjukkan perbedaan signifikan: pada Sekolah-1 diperoleh MAE 35,39, MSE 9.925,15, dan RMSE 99,6; sedangkan pada Sekolah-2 diperoleh MAE 62,77, MSE 25.665,69, dan RMSE 160,2. Hasil ini membuktikan bahwa model lebih akurat dalam memprediksi tren polutan pada Sekolah-1 dibandingkan Sekolah-2. Penelitian ini dapat menjadi tahap awal dalam kontribusi pembuatan sistem pemantauan kualitas udara dalam ruangan dalam bentuk dashboard semi interaktif berbasis Streamlit.

Kata Kunci : Kualitas udara dalam ruangan, LSTM, *Data Analytics*, Streamlit.

ABSTRACT

Application of Data Analytics for Indoor Air Quality Prediction and Visualization Using LSTM to Support Decision-Making

By

Muhammad Afif Rafi Syaiim

Indoor air quality significantly affects health, comfort, and productivity, especially in enclosed spaces such as classrooms and laboratories. A common challenge lies in the difficulty of continuous monitoring and the need for a predictive system that can support preventive decision-making. This study aims to develop an informative and easy-to-understand system for predicting and visualizing indoor air quality. The methodology follows the *Obtain, Scrub, Explore, Model, and Interpret* (OSEMN) framework for processing historical data, with modeling carried out using the *Long Short-Term Memory* (LSTM) algorithm to predict pollutant trends. Model evaluation was conducted using Mean Absolute Percentage Error (MAPE) as the primary metric, yielding results of 6.90% for School-1 and 11.16% for School-2. Additionally, the Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) indicated significant differences: for School-1, the MAE was 35.39, MSE 9,925.15, and RMSE 99.6; while for School-2, the MAE was 62.77, MSE 25,665.69, and RMSE 160.2. These results demonstrate that the model achieved higher accuracy in predicting pollutant trends at School-1 compared to School-2. This research serves as an initial step toward developing an indoor air quality monitoring system in the form of a semi-interactive dashboard based on Streamlit.

Keywords: Indoor air quality, LSTM, *Data Analytics*, Streamlit.

Judul Skripsi

: **PENERAPAN *DATA ANALYTICS* UNTUK
PREDIKSI DAN VISUALISASI KUALITAS
UDARA DALAM RUANGAN BERBASIS LSTM
DALAM MENDUKUNG PENGAMBILAN
KEPUTUSAN**

Nama Mahasiswa

: **Muhammad Afif Rafi Syaiim**

Nomor Pokok Mahasiswa

: **2115061096**

Program Studi

: **S1 Teknik Informatika**

Jurusan

: **Teknik Elektro**

Fakultas

: **Teknik**

MENYETUJUI

1. Komisi Pembimbing

Pembimbing Utama

Pembimbing Pendamping


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


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

NIP. 19681207 199703 1 006

NIP. 19880709 201903 2 015

2. Mengetahui

Ketua Jurusan

Ketua Program Studi Teknik

Teknik Elektro

Informatika


Herlinawati, S.T., M.T.


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

NIP. 197103141999032001

NIP. 197312262000122001

MENGESAHKAN**1. Tim Penguji****Ketua****: Dr. 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 : 18 September 2025**

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini, menyatakan dengan sebenarnya bahwa skripsi saya yang berjudul “ Penerapan *Data Analytics* untuk Prediksi dan Visualisasi Kualitas Udara Dalam Ruangan Berbasis LSTM dalam Mendukung Pengambilan Keputusan” merupakan hasil karya saya sendiri. Semua hasil yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila di kemudian hari terbukti bahwa skripsi saya ini merupakan salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan hukum atau akademik yang berlaku.

Bandar Lampung, 18 September 2025
Pernyataan



Muhammad Afif Rafi Syaiim
2115061096

RIWAYAT HIDUP



Penulis dilahirkan di Bandar Lampung pada tanggal 15 Juli 2003 sebagai anak kedua dari pasangan Bapak Fahrurrozi Syaiim dan Ibu Evi Linawati. Penulis memulai pendidikan di TK Al-Hikmah pada tahun 2008, kemudian melanjutkan SD Negeri 1 Sukarame dan lulus pada tahun 2015. Pendidikan menengah pertama diselesaikan di MTS Negeri 2 Bandar Lampung pada tahun 2018, dan pendidikan menengah atas di MAN 1 Bandar Lampung yang diselesaikan pada tahun 2021. Selanjutnya, penulis melanjutkan pendidikan tinggi di Universitas Lampung, Program Studi S1 Teknik Informatika, Jurusan Teknik Elektro, Fakultas Teknik, melalui jalur SBMPTN (Seleksi Bersama Masuk Perguruan Tinggi Negeri). Selama menjalani masa perkuliahan, penulis aktif berpartisipasi dalam berbagai kegiatan, antara lain:

1. Berperan sebagai Asisten Praktikum di Laboratorium Teknik Komputer Universitas Lampung pada tahun 2024 hingga 2025.
2. Menjadi anggota Himpunan Mahasiswa Teknik Elektro (HIMATRO) pada periode 2021–2023
3. Mengikuti program Merdeka Belajar Kampus Merdeka (MBKM) Studi Independen sebagai Data Analyst di PT GreatEdu Global Mahardika.
4. Mengikuti Penelitian MBKM 2024 dengan judul “Sistem Cerdas Rekomendasi Tanaman Perkebunan Berbasis Machine Learning Sebagai Adaptasi Perubahan Iklim” dengan penempatan pada bagian AI.

Selain itu penulis juga melaksanakan Kuliah Kerja Nyata (KKN) di Desa Tiuh Tohou, Kecamatan Menggala, Kabupaten Tulang Bawang, Provinsi Lampung.

MOTTO

هَلْ جَزَاءُ الْإِحْسَنِ إِلَّا الْإِحْسَنُ

“Tidak ada balasan untuk kebaikan selain kebaikan (pula)”

(QS. Ar-Rahman: 60)

“Berbuatlah baik tanpa mengharap balasan, agar kebaikan tersebar dan dunia menjadi tempat yang lebih baik.”

(Anonymus)

PERSEMBAHAN

Alhamdulillahirabbil‘alamin

Segala puji bagi Allah SWT yang telah melimpahkan rahmat dan karunia-Nya sehingga skripsi ini telah diselesaikan dengan baik. Shalawat serta salam semoga senantiasa tercurah kepada Nabi Muhammad SAW, teladan sepanjang masa.

Kupersembahkan Skripsi Ini Kepada

Kedua orang tuaku tercinta, Bapak Fahrurrozi Syaiim dan Ibu Evi Linawati, yang tak pernah lelah melangitkan doa-doa terbaik sepanjang hidup mereka untukku serta memberikan pelajaran yang terpenting dalam hidup ini. Terima kasih sebesar-besarnya atas kasih sayang, dukungan, dan pengorbanan yang tak ternilai, yang tak akan pernah mampu terbalaskan. Semoga Allah senantiasa memberikan perlindungan dunia dan akhirat, Aamiin.

Untuk Kakakku, Salsabila Rafi'Syaiim, terima kasih telah menjadi penyemangat serta, pendengar setiap cerita, dan pemberi ucapan selamat di setiap langkah pencapaianku. Semoga Allah SWT senantiasa melimpahkan rahmat, berkah, dan perlindungan-Nya kepada kita semua. Aamiin.

Seluruh Dosen, segenap civitas, serta rekan rekan seperjuangan Teknik Informatika Universitas Lampung, atas motivasi, diskusi, dan kebersamaan yang menjadikan perjalanan akademik ini penuh makna.

SANWACANA

Alhamdulillahirabbil'alamin, Penulis mengucapkan puji syukur atas kehadiran Allah SWT, Tuhan Yang Maha Esa, atas segala rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi berjudul “Penerapan Data Analytics untuk Prediksi dan Visualisasi Kualitas Udara Dalam Ruangan Berbasis LSTM dalam Mendukung Pengambilan Keputusan”.

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik pada Program Studi S1 Teknik Informatika, Fakultas Teknik, Universitas Lampung. Proses penyusunan skripsi ini berlangsung selama kurang lebih lima bulan, berbekal pengalaman dan pembelajaran yang telah penulis tempuh selama empat tahun masa perkuliahan.

Penulis menyadari bahwa penyelesaian skripsi ini tidak terlepas dari dukungan bimbingan, dan motivasi dari berbagai pihak. Oleh karena itu, dengan tulus hati, penulis mengucapkan terimakasih 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 sekaligus penguji, atas saran, masukan, dan ilmu yang telah diberikan demi perbaikan penelitian ini.
4. Bapak Ir. M. Komarudin, S.T., M.T., selaku Pembimbing Utama, atas bimbingan, bantuan, dan arahan yang sangat berarti selama proses penelitian hingga terselesaikan dengan baik.
5. Ibu Ir.Titin Yulianti, S.T., M.Eng selaku Pembimbing Pendamping, atas masukan dan arahan yang diberikan dalam penyusunan penelitian ini

- sekaligus sebagai Kepala Laboratorium atas dukungan dan fasilitas yang diberikan serta ilmu dan pengetahuan yang telah diberikan kepada penulis.
6. Terutama dan paling utama kepada kedua orang tua, dan keluarga, atas doa, dukungan, semangat, dan bantuan materil yang tiada henti hingga skripsi ini dapat terselesaikan dengan baik.
 7. Seluruh Asisten Laboratorium Teknik Komputer angkatan 2021, atas kerja sama dan kebersamaan yang luar biasa.
 8. Seluruh mahasiswa Teknik Informatika Universitas Lampung atas perjalanan dan kebersamaannya selama ini.
 9. Seluruh civitas akademik Jurusan Teknik Elektro Universitas Lampung, atas segala bantuan administratif dan dukungan selama proses penelitian.
 10. Dan kepada semua pihak yang tidak dapat disebutkan satu per satu, terima kasih atas kontribusi, baik secara langsung maupun tidak langsung, dalam penyelesaian skripsi ini.

Penelitian ini masih jauh dari kata sempurna, dikarenakan terbatasnya pengalaman dan pengetahuan yang dimiliki. Oleh karena itu, diharapkan segala bentuk saran serta masukan bahkan kritik yang membangun dari berbagai pihak. Semoga skripsi ini dapat bermanfaat bagi pada pembaca dan semua pihak khususnya dalam bidang data analytics serta machine learning.

Bandar Lampung, 18 September 2025
Penulis

Muhammad Afif Rafi Syaiim

DAFTAR ISI

	Halaman
ABSTRAK	iii
<i>ABSTRACT</i>	iv
DAFTAR ISI	xiii
DAFTAR GAMBAR	xvi
DAFTAR TABEL	xviii
I. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	3
1.5 Batasan Masalah	4
1.6 Sistematika Penulisan	4
II. TINJAUAN PUSTAKA	6
2.1 Kualitas Udara dalam Ruangan	6
2.2 Polutan Dalam Ruangan	7
2.2.1 Karbon Dioksida (CO ₂)	7
2.2.2 Partikel Debu	8
2.2.3 Suhu dan Kelembapan	9
2.3 <i>Data Analytics</i>	10
2.4 <i>Long Short-Term Memory</i>	14
2.5 <i>Optimizer</i>	20

2.6	Metode OSEMN.....	20
2.7	Visualisasi Data Eksploratif	21
2.8	Interpolarsi Data	25
2.9	Evaluasi Model.....	26
2.10	Simpangan Baku (<i>Standard Deviation</i>).....	30
2.11	Python	31
2.12	Streamlit.....	31
2.13	Google Colaboratory.....	32
2.14	Penelitian Terkait.....	32
III.	METODOLOGI PENELITIAN.....	39
3.1	Waktu dan Tempat	39
3.2	Alat dan Bahan	39
3.2.1	Alat Penelitian.....	40
3.3	Tahapan Penelitian	41
3.3.1.	<i>Obtain</i>	43
3.3.2.	<i>Scrub</i>	47
3.3.3.	<i>Explore</i>	49
3.3.4.	<i>Modelling</i>	53
3.3.6.	<i>Interpret</i>	56
IV.	HASIL DAN PEMBAHASAN.....	59
4.1	Obtain	59
4.1.1	Deskripsi Data.....	59
4.1.2	Validasi Data.....	60
4.2	Scrub.....	63
4.2.1	Filter Data	63
4.2.2	Perubahan Timestamp.....	65

4.2.3	Resampling Data	66
4.2.4	<i>Clipping data</i>	68
4.2.5	Penanganan Nilai Hilang (NaN) dan Nol.....	68
4.3	Explore	73
4.3.1	Perbandingan Data	73
4.3.2	Analisis Deskriptif	74
4.3.3	Perbandingan Suhu Berdasarkan Jenis	76
4.3.4	Hubungan CO ₂ Dengan Tipe Kelas.....	78
4.3.5	CO ₂ Harian Sekolah-1 dan Sekolah-2	80
4.3.6	Penyebaran <i>Particulate Matter</i> (PM).....	82
4.3.7	Korelasi Kolom	83
4.3.8	Distribusi Data	85
4.4	Modeling	93
4.4.1.	Pelatihan Model	93
4.4.2.	Evaluasi Model	99
4.5	Interpret	102
4.5.1.	Struktur <i>Dashboard</i>	104
4.5.2.	Tampilan <i>Dashboard</i>	108
V.	KESIMPULAN DAN SARAN.....	116
5.1.	Kesimpulan.....	116
5.2.	Saran.....	117
	DAFTAR PUSTAKA	119
	LAMPIRAN.....	125

DAFTAR GAMBAR

	Halaman
Gambar 1 Penyebaran <i>Data Analytics</i> [17].....	10
Gambar 2 Kategori Machine Learning (Sumber: [24]).....	15
Gambar 3 Arsitektur RNN (Sumber : [27])	16
Gambar 4 Arsitektur LSTM (Sumber : [27])	17
Gambar 5 Metode OSEMN (Sumber: [31]).....	21
Gambar 6 Contoh Heat Map (Sumber: [32])	22
Gambar 7 Contoh Grafik Linear (Sumber: [33])	23
Gambar 8 Contoh Scatter plot (Sumber: [8]).....	25
Gambar 9. Flowchart OSEMN.....	42
Gambar 10 Lokasi ruang kelas dan infrastruktur sekolah (Sumber : [50]).....	44
Gambar 11 Arsitektur Pengerjaan.....	45
Gambar 12 Hasil Proses Obtain	46
Gambar 13 Hasil dari proses scrub	49
Gambar 14. Korelasi Setiap Kolom	51
Gambar 15 Hasil Proses Explore	53
Gambar 16. Mock Up Dashboard Kualitas Udara	57
Gambar 17 Filter Data Per Sekolah	63
Gambar 18 Filter Data Per Perangkat Sensor	64
Gambar 19 Perubahan Timestamp	65
Gambar 20 Resampling Data	67
Gambar 21 Clipping Data	68
Gambar 22 Penanganan nilai NaN dan 0	69
Gambar 23 Interpolasi Backfill & Forward Fill	70
Gambar 24 Suhu Berdasarkan Jenis Kelas Sekolah-1	76
Gambar 25 Suhu Berdasarkan Jenis Kelas Sekolah-2	77
Gambar 26 Distribusi CO2 Sekolah-1	78
Gambar 27 Distribusi CO2 Sekolah-2	79

Gambar 28 Rata Rata CO2 Per Jam.....	81
Gambar 29 Variasi Harian Konsentrasi PM1, 2.5, dan 10 pada Interval 1 Jam ...	83
Gambar 30 Heatmap Korelasi.....	84
Gambar 31 Distribusi Kolom Suhu Berdasarkan Frekuensi.....	85
Gambar 32 Distribusi Kolom RH Berdasarkan Frekuensi.....	86
Gambar 33 Distribusi Kolom CO2 Berdasarkan Frekuensi.....	87
Gambar 34 Distribusi Kolom PM1.0 Berdasarkan Frekuensi	88
Gambar 35 Distribusi Kolom PM2.5 Berdasarkan Frekuensi	89
Gambar 36 Distribusi Kolom PM10 Berdasarkan Frekuensi	90
Gambar 37 Distribusi pengambilan data Sekolah-1	91
Gambar 38 Distribusi pengambilan data Sekolah-2	92
Gambar 39 Pembuatan Df_Model	93
Gambar 40 Sorting Bersdarkan Ds	94
Gambar 41 Split Dataset	94
Gambar 42 Normalisasi Data Model.....	95
Gambar 43 Fungsi Create Sequence	95
Gambar 44 Split LSTM.....	96
Gambar 45 Arsitektur LSTM	96
Gambar 46 Pelatihan Model LSTM.....	98
Gambar 47 Prediksi Lstm	98
Gambar 48 Hasil Prediksi LSTM.....	101
Gambar 49 Perubahan Mockup.....	103
Gambar 50 Load Data Streamlit	104
Gambar 51 Pemanggilan Dataset	105
Gambar 52 Sidebar Filter Dashboard.....	106
Gambar 53 Deklarasi Tab	107
Gambar 54 Keseluruhan Dashboard	108
Gambar 55 Tab Time Series Analysis	110
Gambar 56 Temperature & CO2 Distribution	112
Gambar 57 Tab Statistics Data.....	114

DAFTAR TABEL

	Halaman
Tabel 1 Standart Kualitas Udara dalam Ruangan	6
Tabel 2 Satuan Pengukuran CO2 [12]	8
Tabel 3 Satuan Pengukuran PM [14]	8
Tabel 4 Ketegori Data Analytics[20]	11
Tabel 5 Perbandingan Tiap Metrik Evaluasi.....	29
Tabel 6 Penelitian Terkait	34
Tabel 7 Aktivitas Penelitian	39
Tabel 8 Hardware	40
Tabel 9 Software	40
Tabel 10 Struktur Dataset.....	43
Tabel 11 Kolom data yang berisi Null	48
Tabel 12 Contoh analisis deskriptif.....	50
Tabel 13 Kolom yang digunakan	61
Tabel 14 Dictionary Sensor per Ruang Kelas pada Sekolah.....	64
Tabel 15 Perbandingan Format Waktu	66
Tabel 16 Perbandingan resampling data	67
Tabel 17 Evaluasi data hilang sekolah-1.....	71
Tabel 18 Evaluasi data hilang sekolah-2.....	71
Tabel 19 Dataset yang dilanjutkan ke proses	72
Tabel 20 Perbandingan Data	73
Tabel 21 Ringkasan Analisis Deskriptif	74
Tabel 22 Evaluasi Model LSTM Sekolah-1.....	99
Tabel 23 Evaluasi Model LSTM Sekolah-2.....	99

I. PENDAHULUAN

1.1 Latar Belakang

Kualitas udara dalam ruangan menjadi salah satu aspek penting yang kini semakin mendapat perhatian, terutama setelah pandemi COVID-19. Sebelumnya, kesadaran mengenai kualitas udara dalam ruangan masih terbatas, namun pandemi telah meningkatkan pemahaman global akan dampaknya terhadap kesehatan. Ruangan tertutup, seperti kelas, laboratorium, dan kantor, memiliki potensi tinggi dalam menyebarkan polutan udara serta mikroorganisme patogen yang dapat memengaruhi kesehatan penghuninya. Penelitian menunjukkan bahwa ventilasi yang buruk dan akumulasi polutan udara seperti karbon dioksida (CO₂), partikel halus (PM_{2.5}, PM₁₀), serta senyawa organik volatil (VOC) dapat meningkatkan risiko penyakit pernapasan dan penyakit menular [1], [2]. Selain itu, tingginya konsentrasi CO₂ dalam ruangan juga dikaitkan dengan penurunan fungsi kognitif, menurunnya produktivitas, dan gangguan kenyamanan penghuni. Kesadaran akan pentingnya kualitas udara dalam ruangan kini semakin meningkat, menjadikannya aspek krusial yang perlu diperhatikan untuk menciptakan lingkungan yang sehat dan mendukung produktivitas penghuni [3].

Mengingat sifat dinamis dari kualitas udara dalam ruangan, pemantauan manual sering kali tidak efektif dalam mendeteksi perubahan polutan secara real-time. Dengan perkembangan teknologi sensor dan *Internet of Things* (IoT), kini data kualitas udara dapat diperoleh dalam jumlah besar dan dalam interval waktu yang lebih sering [4], [5]. Data yang dihasilkan cenderung kompleks dan sulit untuk diinterpretasikan secara langsung tanpa teknik analisis dan visualisasi data yang tepat. Oleh karena itu, diperlukan metode berbasis *Data Analytics* untuk mengolah data kualitas udara, menemukan pola tren yang berulang, serta menyajikan informasi dalam format yang lebih intuitif guna mendukung pengambilan keputusan.

Perkembangan visualisasi data juga memberikan peluang dalam menyajikan informasi kualitas udara secara lebih interaktif dan informatif [6]. *Dashboard* memungkinkan pemangku kepentingan, seperti manajer gedung atau tenaga kesehatan, untuk memantau kondisi ruangan secara efisien dan membuat keputusan berdasarkan data yang akurat. Dengan adanya *dashboard*, pengguna dapat segera mengetahui apakah kondisi udara di suatu ruangan memenuhi standar kelayakan untuk aktivitas tertentu, baik untuk tinggal, bekerja, maupun proses belajar mengajar [7]. Hal ini membantu dalam pengambilan keputusan yang cepat dan tepat, seperti peningkatan ventilasi, pengurangan kepadatan penghuni, atau tindakan mitigasi lainnya untuk menjaga kesehatan dan kenyamanan.

Dalam penelitian ini, visualisasi data akan dilakukan menggunakan Streamlit, sebuah *framework* berbasis Python yang memungkinkan pembuatan *dashboard* semi interaktif untuk menampilkan hasil prediksi kualitas udara dalam ruangan. Melalui grafik linear, pengguna dapat memahami pola perubahan kualitas udara dari waktu ke waktu, sedangkan *scatter plot* membantu dalam melihat hubungan antar variabel seperti CO₂ dan kelembaban. Selain itu, heatmap akan digunakan untuk menganalisis korelasi antara parameter kualitas udara guna memberikan wawasan lebih mendalam mengenai faktor-faktor yang memengaruhi kondisi lingkungan dalam ruangan [8].

Penelitian akan menerapkan *data analytics* dalam pemodelan dan visualisasi eksploratif kualitas udara dalam ruangan menggunakan algoritma *Long Short-Term Memory* (LSTM) serta menampilkan hasilnya dalam *dashboard* berbasis Streamlit. LSTM dipilih karena kemampuannya dalam menganalisis pola data berbasis time series serta membuat prediksi berdasarkan tren historis. Dengan adanya sistem ini, pengguna dapat memahami pola perubahan kualitas udara, memprediksi kemungkinan peningkatan polutan, serta mengambil langkah-langkah pencegahan secara proaktif.

1.2 Rumusan Masalah

Adapun rumusan masalah dari penelitian ini adalah sebagai berikut:

1. Bagaimana penggunaan *data analytics* untuk menganalisis dan mengolah data kualitas udara dalam ruangan guna mengidentifikasi pola tren serta faktor-faktor yang memengaruhi kualitas udara?
2. Bagaimana mengembangkan model prediksi kualitas udara dalam ruangan menggunakan algoritma *Long Short-Term Memory* (LSTM) dan berdasarkan data historis yang tersedia?
3. Bagaimana menyajikan hasil prediksi kualitas udara dalam ruangan dalam bentuk visualisasi semi interaktif berbasis Streamlit agar mudah dipahami dan digunakan oleh pemangku kepentingan?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Memproses data kualitas udara dalam ruangan menggunakan *data analytics* untuk mengidentifikasi pola tren yang memengaruhi kualitas udara dalam ruangan.
2. Mengembangkan model prediksi kualitas udara dalam ruangan menggunakan algoritma *Long Short-Term Memory* (LSTM) berdasarkan data historis yang tersedia.
3. Menyajikan hasil prediksi kualitas udara dalam bentuk visualisasi semi interaktif berbasis Streamlit agar dapat dengan mudah dipahami oleh pemangku kepentingan.

1.4 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut:

1. Menambah wawasan dan referensi dalam bidang *data analytics*, dan visualisasi data untuk analisis kualitas udara dalam ruangan.
2. Membantu pemangku kepentingan untuk memantau dan memprediksi kualitas udara dalam ruangan.
3. Mempermudah guru dan siswa dalam memantau dan mengakses informasi kualitas udara dalam ruangan yang digunakan.

1.5 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut.

1. Data yang digunakan dalam model adalah data historis, bukan data real-time.
2. *Dashboard* semi interaktif yang dihasilkan merupakan *dashboard* semi interaktif half-duplex.

1.6 Sistematika Penulisan

Adapun sistematika penulisan dari penelitian ini sebagai berikut :

BAB I. PENDAHULUAN

Bab ini menguraikan latar belakang, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan. Bab ini membahas pentingnya kualitas udara dalam ruangan serta bagaimana *data analytics* dan visualisasi dapat membantu dalam pemantauan dan pengambilan keputusan. Dengan tujuan untuk menganalisis pola tren kualitas udara berdasarkan data historis dan menyajikan hasilnya dalam *dashboard* semi interaktif berbasis Streamlit.

BAB II. TINJAUAN PUSTAKA

Bab ini membahas teori dan konsep yang relevan dengan penelitian, seperti definisi dan komponen kualitas udara dalam ruangan, variabel polutan, serta *Long Short-Term Memory*. Bab ini juga mencakup penjelasan mendalam tentang metode visualisasi data eksploratif, seperti *heatmap*, grafik linier, dan *scatter plot*, yang digunakan untuk memvisualisasikan hasil analisis data.

BAB III. METODOLOGI PENELITIAN

Bab ini menjelaskan tahapan-tahapan dalam penelitian, mulai dari proses *data analytics* seperti pengumpulan data kualitas udara dalam ruangan menggunakan data historis, proses preprocessing data, pelatihan model dan evaluasi model.

BAB IV. HASIL DAN PEMBAHASAN

Bab ini berisi hasil yang diperoleh dari model kualitas udara yang telah dibangun. Hasil prediksi dibandingkan dengan data aktual kualitas udara, dan dilakukan analisis terkait akurasi model serta efektivitas visualisasi data dalam mempermudah interpretasi hasil. Pembahasan juga mencakup analisis terhadap variabel-variabel utama yang memengaruhi kualitas udara dalam ruangan, serta bagaimana visualisasi data membantu dalam memahami pola-pola yang muncul. Bab ini juga memberikan rekomendasi bagi pengelola fasilitas atau pihak terkait untuk menggunakan visualisasi analisa data ini dalam mengelola kualitas udara dalam ruangan, seperti melakukan perbaikan ventilasi atau pengurangan aktivitas yang dapat memperburuk kualitas udara.

BAB V. PENUTUP

Bab ini menyajikan kesimpulan dari penelitian yang dilakukan, serta saran untuk pengembangan model prediksi dan *data analytics* yang lebih baik di masa mendatang.

II. TINJAUAN PUSTAKA

2.1 Kualitas Udara dalam Ruangan

Kualitas udara dalam ruangan (*Indoor Air Quality*) merujuk pada kondisi udara di dalam ruang tertutup yang dapat memengaruhi kesehatan dan kenyamanan penghuninya [9]. Kualitas udara dalam ruangan merupakan ukuran terhadap kondisi udara dalam suatu ruangan dan didalamnya mencakup susunan fisik, kimia dan mikrobiologi udara di dalam dan di sekitar bangunan, terutama yang berkaitan dengan kesehatan dan kenyamanan penghuni bangunan [10].

Kualitas udara yang buruk dapat menyebabkan berbagai dampak kesehatan, termasuk gangguan pernapasan, iritasi mata, hingga penyakit kronis seperti asma dan kanker paru-paru. Faktor-faktor seperti ventilasi yang tidak memadai, penggunaan bahan bangunan yang mengandung zat berbahaya, serta akumulasi polutan seperti karbon dioksida (CO_2), formaldehida (CH_2O), partikulat halus ($\text{PM}_{2.5}$ dan PM_{10}), nitrogen dioksida (NO_2), dan karbon monoksida (CO) berkontribusi terhadap menurunnya kualitas udara dalam ruangan [10], [11].

Untuk memastikan kesehatan dan kenyamanan penghuni, berbagai organisasi internasional telah menetapkan standar kualitas udara dalam ruangan. Berikut adalah batas aman konsentrasi polutan berdasarkan pedoman WHO dan penelitian terbaru [10] :

Tabel 1 Standart Kualitas Udara dalam Ruangan

Polutan	Batas konsentrasi yang direkmendasikan
Karbon Dioksida (CO_2)	<1000 ppm (baik), 1000–1500 ppm (moderat), >1500 ppm (buruk)
Karbon Monoksida (CO)	100 mg/m^3 (15 menit), 35 mg/m^3 (1 jam), 10 mg/m^3 (8 jam), 7 mg/m^3 (24 jam)
Formaldehida (CH_2O)	100 $\mu\text{g}/\text{m}^3$ (30 menit)

Polutan	Batas konsentrasi yang direkomendasikan
Nitrogen Dioksida (NO ₂)	200 µg/m ³ (1 jam), 40 µg/m ³ (tahunan)
Partikulat Halus (PM _{2.5})	25 µg/m ³ (24 jam), 10 µg/m ³ (tahunan)
Radon (Rn)	Maksimal 100 Bq/m ³ , batas maksimum 300 Bq/m ³ di daerah tertentu
Suhu Ruangan (T)	18–24°C
Kelembaban Relatif (RH)	40–60%

Selain parameter polutan, kualitas udara dalam ruangan juga sangat bergantung pada kondisi ventilasi, suhu, dan kelembaban. WHO merekomendasikan suhu dalam ruangan berada pada 18–24°C untuk kenyamanan termal, sementara kelembaban relatif 40–60% dianggap optimal untuk mencegah pertumbuhan jamur dan tungau debu [11].

2.2 Polutan Dalam Ruangan

Menurut penelitian, polusi udara dalam ruangan dapat berasal dari berbagai sumber, termasuk aktivitas manusia, bahan bangunan, dan ventilasi yang buruk [1]. Parameter utama yang digunakan untuk mengevaluasi IAQ meliputi karbon dioksida (CO₂), materi partikulat (PM_{2.5}, PM₁₀), suhu, dan kelembaban relatif [10]

2.2.1 Karbon Dioksida (CO₂)

Karbon dioksida adalah gas yang tidak berwarna dan tidak berbau yang secara alami ada di atmosfer sebagai bagian dari siklus karbon. CO₂ biasanya dihasilkan oleh proses pernapasan manusia dan hewan serta pembakaran bahan bakar fosil. Dalam konteks kualitas udara dalam ruangan, CO₂ berperan sebagai indikator ventilasi. Konsentrasi tinggi CO₂ di dalam ruangan menunjukkan buruknya sirkulasi udara, yang dapat mengakibatkan penurunan kinerja kognitif dan kelelahan pada penghuni ruangan.

Paparan CO₂ dalam jumlah tinggi dapat berdampak pada kesehatan manusia, termasuk penurunan fungsi kognitif, sakit kepala, kelelahan, dan kesulitan berkonsentrasi. Sebuah studi menunjukkan bahwa peningkatan CO₂ di atas 1000 ppm dapat menurunkan kinerja kognitif hingga 15%, sementara di atas 2500 ppm

dapat menurunkan fungsi pengambilan keputusan hingga 50% [10]. CO₂ diukur dalam ppm (parts per million), dengan standar berikut:

Tabel 2 Satuan Pengukuran CO₂ [12]

Konsentrasi CO ₂ (ppm)	Dampak pada kesehatan dan kenyamanan
400-600 ppm	Level normal di udara luar
600-1000 ppm	Baik, menandakan ventilasi cukup
1000-1500 ppm	Moderat, dapat menyebabkan kelelahan dan gangguan fokus
>1500 ppm	Buruk, berisiko menyebabkan kantuk, sakit kepala, dan kinerja kognitif menurun

2.2.2 Partikel Debu

Partikel debu terdiri dari partikel-partikel kecil padat atau cair yang tersuspensi di udara. Berdasarkan ukuran partikel, partikulat ini dikategorikan menjadi PM₁₀ (partikel dengan diameter kurang dari 10 mikrometer) dan PM_{2.5} (diameter kurang dari 2.5 mikrometer). PM_{2.5} dianggap lebih berbahaya karena partikel yang sangat kecil ini dapat masuk ke dalam sistem pernapasan hingga ke alveoli di paru-paru, menyebabkan gangguan kesehatan serius seperti asma, bronkitis, dan penyakit kardiovaskular. Partikel debu di dalam ruangan dapat berasal dari luar, serta dihasilkan dari aktivitas manusia di dalam ruangan, seperti merokok, penggunaan peralatan rumah tangga, atau pembakaran bahan-bahan tertentu [13].

Sumber partikel debu dalam ruangan bisa berasal dari pembakaran bahan bakar, asap rokok, lilin, aktivitas memasak, penggunaan alat elektronik, serta polusi luar yang masuk melalui ventilasi. PM_{2.5} lebih berbahaya karena ukurannya yang kecil memungkinkan penetrasi lebih dalam ke sistem pernapasan, berpotensi menyebabkan asma, bronkitis, penyakit kardiovaskular, serta peningkatan risiko kematian dini [10]. PM diukur dalam $\mu\text{g}/\text{m}^3$ (mikrogram per meter kubik), dengan standar berikut:

Tabel 3 Satuan Pengukuran PM [14]

Jenis PM	Batas konsentrasi WHO
PM ₁₀	50 $\mu\text{g}/\text{m}^3$ (24 jam), 20 $\mu\text{g}/\text{m}^3$ (tahunan)
PM _{2.5}	25 $\mu\text{g}/\text{m}^3$ (24 jam), 10 $\mu\text{g}/\text{m}^3$ (tahunan)

2.2.3 Suhu dan Kelembapan

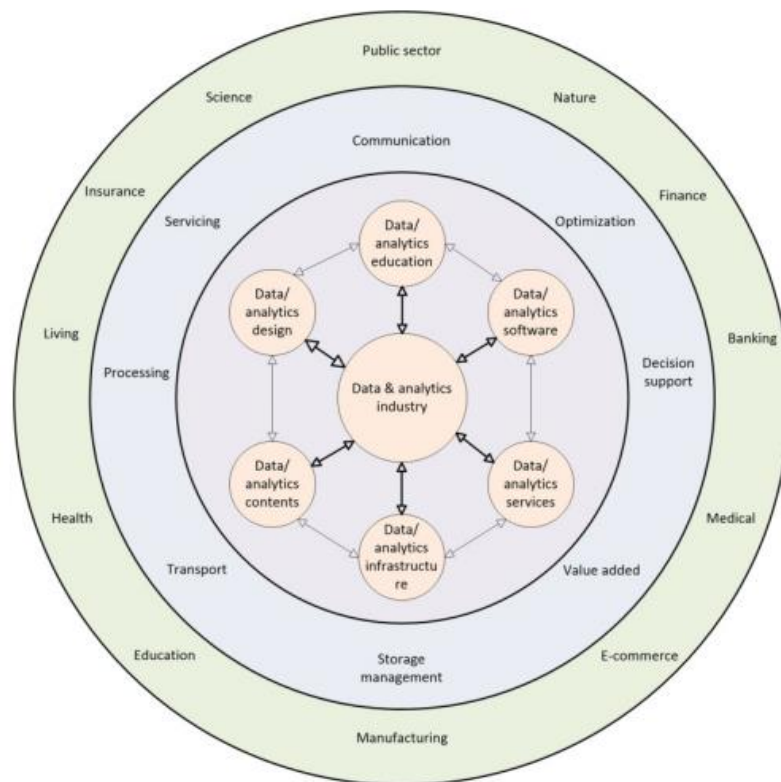
Suhu adalah ukuran derajat panas atau dinginnya suatu lingkungan yang biasanya diukur dalam derajat Celsius ($^{\circ}\text{C}$) atau Fahrenheit ($^{\circ}\text{F}$). Dalam konteks kualitas udara dalam ruangan, suhu yang nyaman sangat penting untuk kenyamanan penghuni dan efisiensi kerja. Suhu yang terlalu tinggi atau terlalu rendah dapat menyebabkan ketidaknyamanan fisik, mengganggu konsentrasi, serta mempengaruhi kesehatan seperti menyebabkan stres termal atau hipotermia [15].

Kelembapan adalah kadar uap air yang terdapat di udara dan biasanya diukur sebagai kelembapan relatif (%RH). Kelembapan relatif adalah persentase jumlah uap air yang ada di udara dibandingkan dengan jumlah maksimum uap air yang bisa ditahan udara pada suhu tertentu. Kelembapan yang terlalu rendah dapat menyebabkan iritasi pada saluran pernapasan dan kulit kering, sedangkan kelembapan yang terlalu tinggi dapat menyebabkan pertumbuhan jamur, bakteri, dan memperburuk kualitas udara [15].

Menurut prinsip termodinamika, udara yang lebih hangat memiliki kemampuan lebih besar untuk menahan uap air dalam bentuk gas. Hal ini disebabkan oleh peningkatan tekanan uap jenuh secara eksponensial seiring kenaikan suhu udara, yang berarti udara hangat dapat menampung uap air lebih banyak tanpa terjadi kondensasi [16]. Ketika suhu udara meningkat dan jumlah uap air tetap sama, tekanan uap jenuh naik sehingga rasio ini (kelembapan relatif) menurun. Kondisi ini menjelaskan mengapa udara hangat sering kali terasa lebih “kering” meski jumlah uap air sebenarnya tidak berubah. Kombinasi suhu dan kelembapan yang tidak seimbang dapat mempengaruhi kenyamanan termal serta kesehatan penghuni, karena kelembapan yang terlalu rendah dapat menyebabkan iritasi, sementara kelembapan tinggi pada suhu hangat dapat memicu pertumbuhan mikroorganisme. Suhu ideal berkisar antara $22\text{--}24^{\circ}\text{C}$ dengan kelembapan relatif $40\text{--}60\%$, yang dapat meningkatkan kinerja dan menurunkan risiko infeksi [15].

2.3 Data Analytics

Pada ekosistem *Data & Analytics Industry*, keempat kategori ini terintegrasi melalui pilar-pilar inti yaitu education, software, services, infrastructure, contents, design dan diwujudkan sebagai fungsi-fungsi operasional seperti processing, storage management, servicing, communication, optimization, decision support, value added yang diaplikasikan di berbagai sektor termasuk pendidikan dan pengelolaan kualitas udara dalam ruangan [17].



Gambar 1 Penyebaran *Data Analytics* [17]

Penelitian ini menitikberatkan pada *predictive analytics* dengan LSTM, namun tetap memerlukan *descriptive & diagnostic analytics* untuk pra-pemrosesan data dan interpretasi awal.

Data analytics merupakan disiplin ilmu yang berfokus pada pengolahan, analisis, dan interpretasi data untuk mendapatkan wawasan yang dapat digunakan dalam pengambilan keputusan [18]. Dalam era digital saat ini, jumlah data yang dihasilkan terus meningkat secara eksponensial, menciptakan tantangan dalam

pemrosesan dan pemanfaatan informasi [19]. Oleh karena itu, teknik *Data analytics* menjadi semakin penting dalam berbagai bidang, termasuk dalam pemantauan dan prediksi kualitas udara dalam ruangan. *Data analytics* memungkinkan analisis terhadap pola dan tren dalam data yang kompleks serta membantu dalam mengidentifikasi faktor-faktor yang berkontribusi terhadap perubahan suatu fenomena.

Secara umum, *data analytics* dapat dikategorikan menjadi empat jenis utama berdasarkan tujuan dan pendekatan analisisnya:

Tabel 4 Kategori *Data Analytics*[20]

Kategori	Deskripsi	Metode yang umum digunakan
<i>Descriptive Analytics</i> (Analisis Deskriptif)	Menganalisis data historis untuk memahami pola dan tren. Menjawab pertanyaan: “Apa yang telah terjadi?”	Statistik deskriptif (mean, median, mode, distribusi data)
<i>Diagnostic Analytics</i> (Analisis Diagnostik)	Menentukan penyebab dari suatu pola atau kejadian dalam data. Menjawab pertanyaan: “Mengapa hal ini terjadi?”	Analisis regresi, clustering, korelasi antar variabel
<i>Predictive Analytics</i> (Analisis Prediksi)	Memprediksi kejadian di masa depan berdasarkan pola dalam data historis. Menjawab pertanyaan: “Apa yang kemungkinan akan terjadi?”	<i>Machine Learning</i> (LSTM, SVM, <i>Random Forest</i> , ARIMA)
<i>Prescriptive Analytics</i> (Analisis Perspektif)	Memberikan rekomendasi tindakan berdasarkan hasil prediksi. Menjawab pertanyaan: “Apa yang sebaiknya dilakukan?”	Sistem pendukung keputusan berbasis AI, optimasi algoritma

Dalam penelitian ini, pendekatan yang digunakan adalah *Predictive Analytics*, dimana model LSTM diterapkan untuk memprediksi kualitas udara dalam ruangan berdasarkan data historis yang tersedia.

Dalam konteks *machine learning*, *data analytics* merupakan bagian dari keseluruhan siklus pengembangan model, seperti proses eksplorasi data, pembersihan, dan seleksi fitur yang relevan. Menurut penelitian terbaru, kualitas data yang buruk dapat menurunkan suatu akurasi model sehingga peran *data engineering* dan *data analytics* sangat penting untuk meningkatkan efisiensi pipeline data dan memastikan kualitas dataset yang optimal sebelum model dilatih [21]. Dalam penelitian ini, pendekatan *Predictive Analytics* digunakan dengan menerapkan *Long Short-Term Memory* (LSTM) untuk memprediksi kualitas udara dalam ruangan berdasarkan data historis. Proses ini melibatkan pengumpulan data, preprocessing, feature selection, model training, dan evaluasi guna memastikan model dapat mengidentifikasi pola tren yang akurat. Selain itu, dengan optimasi *data engineering*, waktu pemrosesan data dapat efektif yang memungkinkan model LSTM bekerja lebih efisien dalam menganalisis data time-series [21].

Dalam proses analisis data kualitas udara, tahap pra-pemrosesan menjadi krusial untuk memastikan kualitas data yang digunakan. Kualitas data adalah aspek krusial yang menentukan keandalan dan efektivitas data sebagai dasar untuk proses pengolahan dan pengambilan keputusan [22]. Kualitas data tidak hanya terkait dengan ketepatan informasi, tetapi juga mencakup beberapa dimensi utama yang perlu diukur dan dikelola secara sistematis agar data dapat memenuhi kebutuhan pengguna. Empat dimensi kualitas data yang paling umum dan signifikan adalah [22]:

- *Consistency* : Mengindikasikan keseragaman data di berbagai sistem atau sumber sehingga tidak terjadi pertentangan atau konflik data. Konsistensi menjaga integritas data dan meningkatkan kepercayaan pengguna.

- *Accuracy* : Menandakan sejauh mana data mencerminkan nilai atau kondisi sebenarnya di dunia nyata. Data yang akurat memungkinkan pengambilan keputusan yang tepat dan mengurangi risiko kesalahan.
- *Completeness* : Mengacu pada sejauh mana data yang dibutuhkan tersedia secara menyeluruh tanpa adanya nilai yang hilang atau kosong. Data yang lengkap penting untuk memastikan informasi yang komprehensif dan valid.
- *Timeliness* : Berkaitan dengan ketersediaan data pada waktu yang dibutuhkan, sehingga data tidak menjadi usang atau tidak relevan. Ketepatan waktu meningkatkan kegunaan data dalam proses pengambilan keputusan yang cepat dan tepat.

Dengan menerapkan dan memastikan keempat dimensi kualitas data tersebut, maka data dapat dikategorikan sebagai data bersih. Data bersih adalah data yang telah melalui proses pembersihan dan validasi sehingga memenuhi standar kualitas yang diperlukan untuk mendukung analisis dan pengambilan keputusan secara optimal [22].

Salah satu teknik yang sering diterapkan adalah clipping, yaitu membatasi nilai-nilai data pada rentang fisik yang valid sesuai karakteristik parameter yang diukur. Untuk konsentrasi CO₂, nilai yang masuk akal biasanya berkisar antara 400 hingga 5000 ppm . Data yang berada di luar batas ini umumnya dianggap sebagai noise atau kesalahan pengukuran sensor yang dapat menurunkan kualitas hasil analisis.

Dengan menerapkan clipping, nilai ekstrim yang tidak realistis dihapus sehingga data menjadi lebih konsisten dan representatif. Hal ini berdampak positif pada tahap analisis selanjutnya seperti visualisasi, identifikasi pola tren, serta pembangunan model prediksi yang lebih akurat. Penelitian sebelumnya juga menegaskan bahwa clipping merupakan bagian dari data cleansing yang efektif untuk meningkatkan performa model machine learning, termasuk dalam konteks

prediksi kualitas udara berbasis Long Short-Term Memory (LSTM) dan teknik analytics lainnya

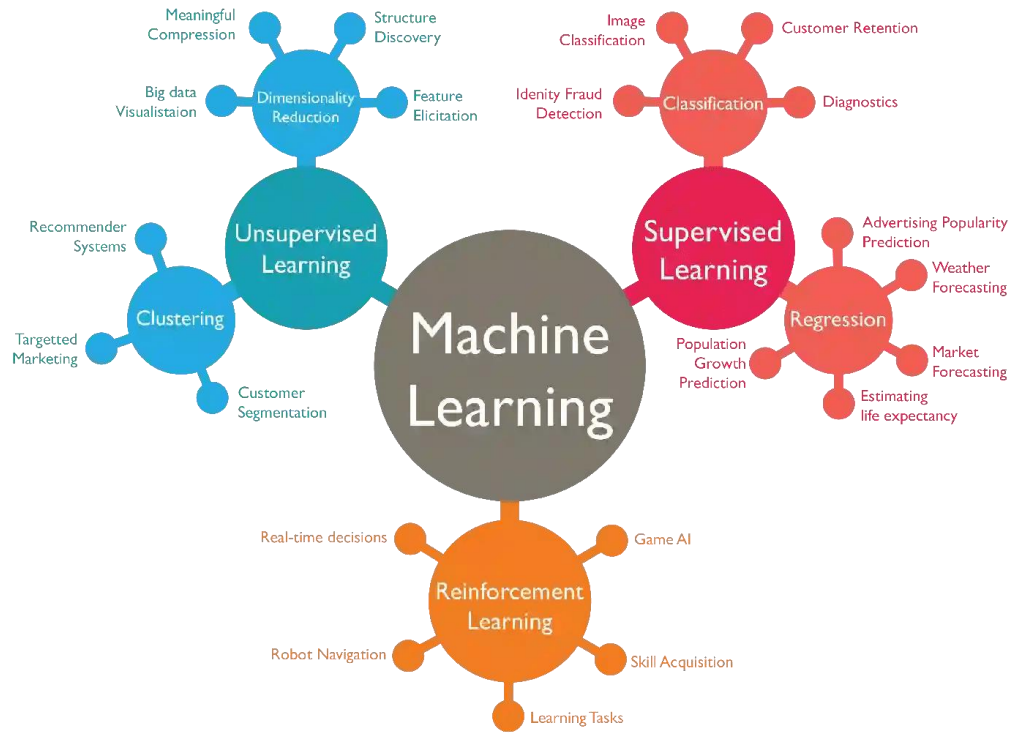
2.4 Long Short-Term Memory

Long Short-Term Memory (LSTM) merupakan jenis *Neural Network* yang dirancang khusus untuk menangani permasalahan sequential data atau deret waktu. LSTM dikembangkan sebagai solusi atas keterbatasan *Recurrent Neural Network* dalam menangani masalah *vanishing gradient*, yaitu ketika jaringan mengalami kesulitan dalam menyimpan dan memproses informasi dalam jangka panjang [23]. Dengan arsitektur berbasis *gating mechanism*, LSTM mampu mempertahankan informasi yang relevan dalam rentang waktu panjang, sehingga sangat cocok untuk analisis data deret waktu seperti prediksi kualitas udara dalam ruangan.

LSTM termasuk dalam kategori *Unsupervised Learning* dan *Supervised Learning*. Secara lebih luas, LSTM termasuk dalam cabang *deep learning*, yang merupakan perkembangan dari *Machine Learning*. *Machine Learning* sendiri adalah subbidang dari kecerdasan buatan yang berfokus pada pengembangan algoritma yang dapat belajar dari data tanpa pemrograman eksplisit [24]. *Machine Learning* berkembang dari disiplin ilmu statistik dan kecerdasan buatan, dengan fokus pada pembuatan model prediksi berdasarkan pola dalam data. Seiring perkembangan teknologi, *Machine Learning* telah menjadi dasar bagi berbagai inovasi, termasuk dalam bidang kesehatan, keuangan, dan lingkungan.

Dalam *Machine Learning*, terdapat tiga kategori utama, yaitu *Supervised Learning*, *Unsupervised Learning*, dan *Reinforcement Learning*. *Supervised Learning* digunakan ketika data memiliki label atau target yang jelas, seperti dalam regresi atau klasifikasi seperti algoritma *Linear Regression*, *Random Forest*, *SVM*, *LSTM*. *Unsupervised Learning*, sebaliknya, digunakan untuk menemukan pola tersembunyi dalam data yang tidak memiliki label, seperti dalam clustering, dimensionality reduction, *K-Means*, *DBSCAN*, *PCA*, *Autoencode*. Sedangkan *Reinforcement Learning* lebih berfokus pada pembelajaran berbasis

umpan balik dari lingkungan untuk mengoptimalkan pengambilan keputusan seperti algoritma Q-Learning, Deep Q-Network (DQN), PPO. Kategori *machine learning* dapat dilihat pada Gambar 2.



Gambar 2 Kategori *Machine Learning* (Sumber: [24])

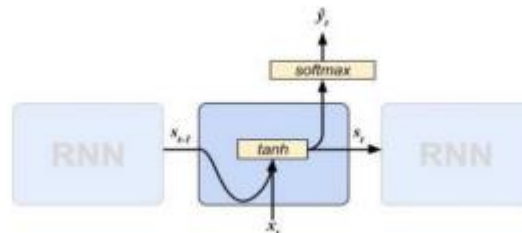
Seiring berkembangnya *Machine Learning*, muncul *Deep Learning*, yang mengandalkan arsitektur jaringan saraf tiruan yang lebih kompleks untuk mengekstrak pola dari data [25]. *Deep Learning* dirancang untuk mengekstrak fitur secara otomatis dari data mentah dengan menggunakan berbagai transformasi non-linear [25].

Deep Learning sangat kuat dalam menangani data yang kompleks dan besar, ia memiliki keterbatasan dalam menangani data dengan pola musiman atau komponen tren jangka panjang yang seringkali ditemukan dalam masalah peramalan deret waktu. Terdapat beberapa model seperti , yang dirancang khusus untuk menangani komponen musiman dan tren jangka panjang pada data deret waktu. Berbeda dengan pendekatan *Deep Learning*, menggunakan metode yang lebih eksplisit dalam menangani fluktuasi musiman dan liburan, menjadikannya

sangat cocok untuk data yang memiliki karakteristik musiman yang jelas, seperti kualitas udara atau data keuangan.

Deep Learning berkembang sebagai solusi untuk mengatasi keterbatasan *Machine Learning* tradisional, yang sering kali membutuhkan *feature engineering* manual. Dengan menggunakan multi-layer architecture, *Deep Learning* mampu menangkap pola data yang lebih kompleks, terutama dalam *computer vision*, *natural language processing*, dan analisis time series [24].

Salah satu bentuk awal Deep Learning adalah Jaringan Saraf Tiruan atau *Artificial Neural Network* (ANN), yang terinspirasi dari cara kerja otak manusia dalam memproses informasi. ANN konvensional memiliki keterbatasan dalam menangani data sekuensial karena tidak dapat mempertahankan informasi dari input sebelumnya secara efektif [26]. Untuk mengatasi keterbatasan ini, dikembangkan *Recurrent Neural Network* (RNN), yang memiliki koneksi berulang sehingga memungkinkan jaringan untuk memiliki "memori" terhadap data sebelumnya seperti yang dapat dilihat pada Gambar 3.

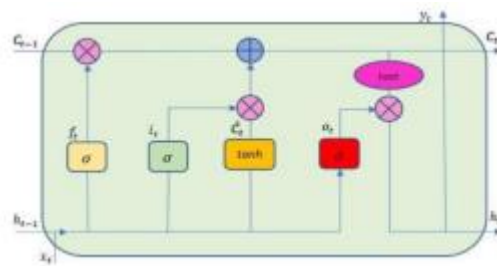


Gambar 3 Arsitektur RNN (Sumber : [27])

Recurrent Neural Network (RNN) merupakan salah satu jenis jaringan saraf tiruan yang dirancang khusus untuk menangani data sekuensial dan time series. Tidak seperti jaringan saraf *feedforward* yang hanya mempertimbangkan input saat ini, RNN memiliki memori internal yang memungkinkan informasi dari langkah waktu sebelumnya digunakan dalam perhitungan saat ini [28]. RNN masih menghadapi kendala dalam menangani ketergantungan jangka panjang akibat *vanishing gradient problem*, yang menyebabkan informasi dari langkah-langkah sebelumnya sulit dipertahankan saat melalui banyak lapisan propagasi. Oleh karena itu, dikembangkan LSTM, yang memperkenalkan unit memori dan *gate*

mechanism untuk secara selektif menyimpan atau membuang informasi, sehingga lebih efektif dalam mempelajari pola dari data sekuensial.

Long Short-Term Memory (LSTM) adalah perkembangan dari *Recurrent Neural Network* yang dirancang untuk menangani masalah *vanishing gradient* pada pemrosesan data sekuensial dan time series. Model ini pertama kali diperkenalkan oleh Hochreiter dan Schmidhuber (1997) dan telah menjadi metode unggulan dalam berbagai aplikasi prediksi jangka panjang, termasuk dalam *forecasting* kualitas udara dalam ruangan [29]. LSTM memiliki cell state sebagai jalur utama yang membawa informasi sepanjang proses, dan beberapa gate yang mengontrol aliran informasi. Berikut penjelasan arsitektur berdasarkan Gambar 4:



Gambar 4 Arsitektur LSTM (Sumber : [27])

1. Forget Gate (Kotak kuning)

Gate ini menentukan informasi mana yang harus dilupakan dari cell state sebelumnya (c_{t-1}) Rumusnya:

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

Keterangan:

- f_t : Forget gate vector pada waktu ttt, berisi nilai 0–1
- σ : Fungsi sigmoid yang menghasilkan nilai antara 0 hingga 1.
- W_f : Bobot (weight matrix) khusus untuk forget gate.
- $[h_{t-1}, x_t]$: Gabungan antara hidden state sebelumnya dan input saat ini.
- b_f : Bias untuk forget gate

2. Input Gate (Kotak hijau)

Gate ini menentukan informasi baru apa yang akan disimpan ke dalam cell state. Rumusnya:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

Keterangan:

- i_t : *Forget gate vector* pada waktu ttt, berisi nilai 0–1
- σ : Fungsi sigmoid yang menghasilkan nilai antara 0 hingga 1.
- W_i : Bobot (*weight matrix*) khusus untuk forget gate.
- $[h_{t-1}, x_t]$: Gabungan antara hidden state sebelumnya dan input saat ini.
- b_i : Bias untuk *forget gate*

3. Candidate Cell State (Kotak oranye)

Ini adalah kandidat nilai baru yang dihasilkan dari tanh. Rumusnya:

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$

Keterangan:

- $\tilde{C}_t$: *Candidate cell state*, informasi baru yang diusulkan untuk ditambahkan ke memori.
- Tanh : Fungsi aktivasi *hyperbolic tangent*, hasil antara -1 hingga 1.
- W_C : Matriks bobot untuk *candidate cell state*.
- $[h_{t-1}, x_t]$: Gabungan hidden state sebelumnya dan input saat ini.
- b_C : Bias untuk candidate cell state.

4. Cell State Update

Cell state yang baru adalah gabungan dari hasil forget gate dan input gate:

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (4)$$

Keterangan:

- C_t : *Cell state* baru pada waktu t .
- f_t : *Forget gate*
- C_{t-1} : *Cell state* sebelumnya
- i_t : *Input gate*
- $\tilde{C}_t$: *Candidate cell state* (informasi baru yang diusulkan).

5. Output Gate (Kotak merah)

Gate ini menentukan bagian dari cell state yang akan dikeluarkan sebagai hidden state berikutnya (h_t)

Rumusnya:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

Keterangan:

o_t : Output gate vector, menentukan bagian cell state yang dipakai untuk hidden state.

σ : Fungsi aktivasi sigmoid.

W_o : Matriks bobot untuk output gate.

$[h_{t-1}, x_t]$: Gabungan hidden state sebelumnya dan input saat ini.

b_o : Bias untuk output gate.

6. Hidden State (Oval pink)

Hasil akhir LSTM di langkah waktu tersebut:

$$h_t = o_t \times \tanh(C_t) \quad (6)$$

Keterangan:

h_t : *Hidden state* baru pada waktu ttt, sekaligus output dari LSTM pada langkah tersebut.

o_t : Nilai dari output gate.

$\tanh(C_t)$: *Cell state* yang ditransformasi dengan fungsi Tanh agar nilainya terjaga antara -1 dan 1.

Dalam konteks penelitian ini, LSTM dan akan digunakan untuk time series forecasting pada dataset kualitas udara dalam ruangan. Dataset ini terdiri dari serangkaian parameter lingkungan seperti suhu, kelembaban, karbon dioksida (CO2), dan partikulat debu (PM2.5), yang semuanya bersifat time-dependent. Karena data tersebut memiliki pola yang berulang dalam interval waktu tertentu, model LSTM dan sangat cocok digunakan untuk menangkap tren dan memprediksi nilai di masa depan berdasarkan data historis yang tersedia. Hal ini menjadikan LSTM dan sebagai alat yang sangat berguna dalam menganalisis dan

meningkatkan kualitas udara dalam ruangan, mendukung pengambilan keputusan berbasis data.

2.5 *Optimizer*

Optimizer adalah algoritma yang digunakan dalam proses pelatihan model machine learning maupun deep learning untuk memperbarui parameter (bobot dan bias) dengan tujuan meminimalkan nilai *loss function*. *Optimizer* bekerja dengan memanfaatkan gradien, yaitu turunan dari fungsi loss terhadap bobot, sehingga arah pembaruan parameter bergerak menuju nilai optimal. Rumus *optimizer* sebagai berikut:

$$w_{\text{baru}} = w_{\text{lama}} - \alpha \cdot \nabla L(w) \quad (7)$$

Keterangan:

$w_{\text{baru}}, w_{\text{lama}}$: Bobot

α : *Learning Rate*

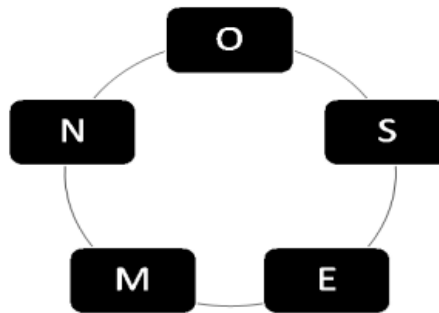
$\nabla L(w)$: Gradien fungsi loss terhadap bobot.

Beberapa *optimizer* yang umum digunakan antara lain *Gradient Descent* yang sederhana namun kurang efisien pada dataset besar, *Stochastic Gradient Descent* (SGD) yang lebih cepat dengan memanfaatkan mini-batch, serta Momentum yang menambahkan komponen inersia untuk mempercepat konvergensi. Selain itu terdapat AdaGrad yang menyesuaikan learning rate adaptif pada tiap parameter, RMSProp yang menjaga learning rate tetap stabil sepanjang pelatihan, dan Adam (*Adaptive Moment Estimation*) yang merupakan kombinasi Momentum dan RMSProp. Adam menjadi salah satu optimizer paling populer karena mampu memberikan konvergensi cepat dan stabil pada berbagai permasalahan *deep learning* [30].

2.6 **Metode OSEM**

Metode OSEM (*Obtain, Scrub, Explore, Model, Interpret*) merupakan pendekatan sistematis yang sering digunakan dalam *data analytics*, terutama dalam bidang *data science* dan *machine learning*. metode OSEM merupakan pendekatan sistematis dalam analisis data yang terdiri dari lima tahap utama: *Obtain* (Mendapatkan data), *Scrub* (Membersihkan data), *Explore* (Menjelajahi

data), Model (Membangun model), dan *Interpret* (Menginterpretasikan hasil) seperti yang dapat dilihat pada Gambar 5.



Gambar 5 Metode OSEM N (Sumber: [31])

Proses ini diawali dengan tahap Obtain, dimana data dikumpulkan dari berbagai sumber seperti database, sensor, atau API. Setelah itu, tahap Scrub dilakukan untuk membersihkan data dari kesalahan, nilai yang hilang, atau inkonsistensi sehingga siap untuk dianalisis. Selanjutnya, tahap Explore berfokus pada eksplorasi data menggunakan teknik visualisasi untuk mengidentifikasi pola, tren, dan outlier yang mungkin mempengaruhi hasil analisis. Setelah eksplorasi selesai, tahap Model dilakukan dengan membangun model prediktif atau analitis, seperti penggunaan algoritma *machine learning* untuk menemukan hubungan dalam data. Terakhir, tahap Interpret bertujuan untuk menginterpretasikan hasil model agar dapat digunakan dalam pengambilan keputusan yang lebih baik. Dengan pendekatan ini, metode OSEM N memberikan struktur yang jelas dan sistematis dalam pengolahan data, memastikan setiap langkah dilakukan dengan optimal sebelum melanjutkan ke tahap berikutnya [31].

2.7 Visualisasi Data Eksploratif

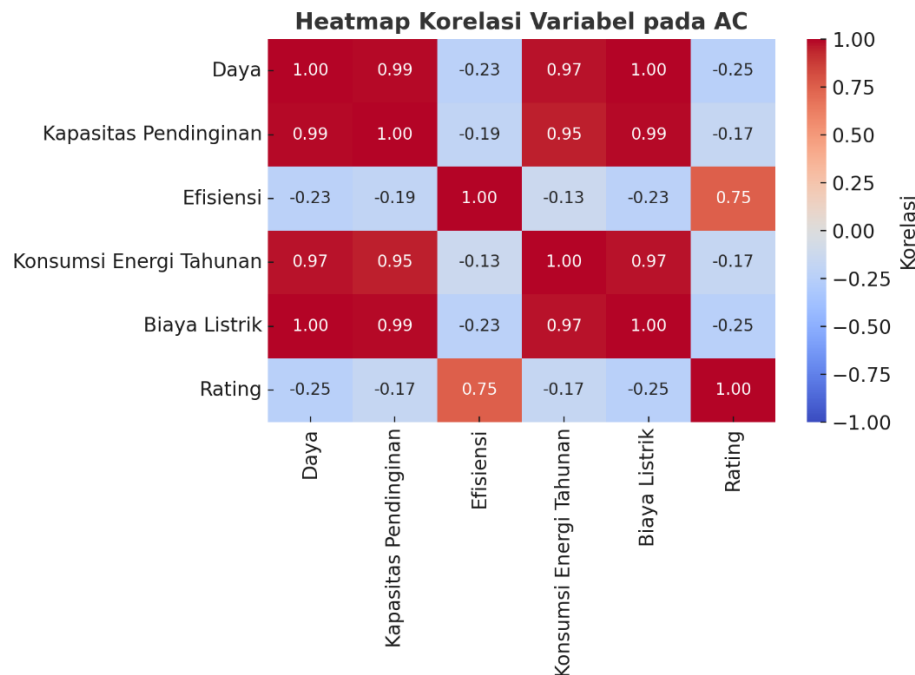
Visualisasi data eksploratif (Exploratory Data Visualization) adalah proses menggunakan grafik dan plot untuk memahami struktur, pola, tren, dan anomali dalam data sebelum melakukan analisis statistik formal atau pemodelan. Visualisasi data eksploratif merupakan teknik yang digunakan untuk menyajikan data dalam format yang mudah dipahami, sehingga memfasilitasi interpretasi dan analisis [8]. Teknik ini membantu dalam memahami distribusi data,

mengidentifikasi tren, serta mendukung pengambilan keputusan berbasis data selain itu visualisasi eksploratif berfungsi untuk mengeksplorasi data secara intuitif, mengidentifikasi hubungan antar variabel, mendeteksi outlier, dan merumuskan hipotesis awal yang dapat diuji lebih lanjut.

Beberapa teknik visualisasi yang umum digunakan dalam penelitian kualitas udara meliputi:

A. *Heatmap*

Heatmap adalah teknik visualisasi yang menggunakan warna untuk merepresentasikan intensitas suatu variabel. Teknik ini digunakan untuk menggambarkan intensitas polutan dalam ruang tertentu, memberikan gambaran visual yang jelas tentang area dengan kualitas udara buruk. Dalam konteks kualitas udara, *heatmap* dapat digunakan untuk menunjukkan distribusi polutan di dalam ruangan. Warna merah dapat menunjukkan konsentrasi tinggi dari gas polutan sementara warna biru dapat menunjukkan daerah dengan kualitas udara lebih baik.

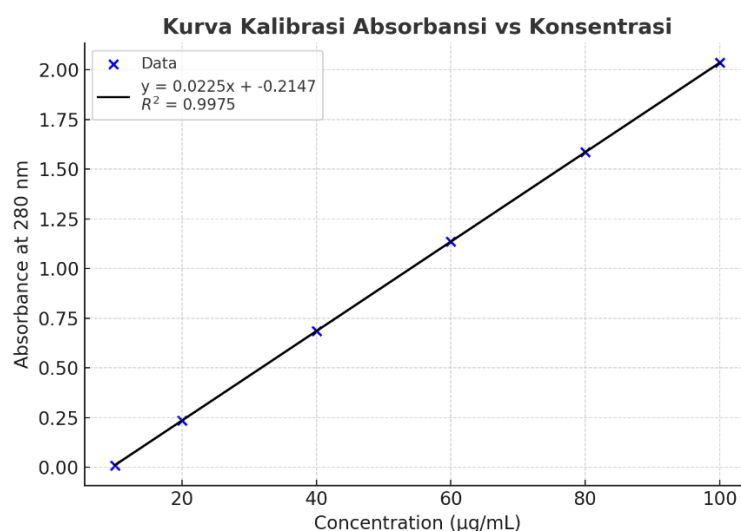


Gambar 6 Contoh Heat Map (Sumber: [32])

Penggunaan *heatmap* sangat efektif untuk analisis data spasial dan temporal. Sebagai contoh, *heatmap* kualitas udara dapat diterapkan dalam pemetaan polusi udara di gedung perkantoran, dimana setiap ruangan atau lantai ditandai berdasarkan tingkat polusinya. *Heatmap* sangat membantu dalam mengidentifikasi pola yang sulit dikenali hanya dengan melihat angka atau grafik biasa [8]. Selain itu, *heatmap* juga sering digunakan dalam analisis tren waktu untuk melihat bagaimana polutan berubah sepanjang hari, minggu, atau bulan. Misalnya, *heatmap* dapat menunjukkan fluktuasi kadar CO₂ selama jam kerja, membantu dalam pengambilan keputusan terkait ventilasi dan tata letak ruangan agar kualitas udara lebih optimal.

B. Grafik Linear

Grafik linear digunakan untuk menampilkan perubahan suatu variabel secara berkelanjutan dalam rentang waktu tertentu. Teknik ini sangat berguna dalam analisis tren data time series, seperti pemantauan perubahan kualitas udara dalam ruangan berdasarkan waktu. Misalnya, grafik linear dapat menunjukkan bagaimana kadar CO₂ atau PM_{2.5} berubah sepanjang hari, membantu mengidentifikasi pola tertentu seperti peningkatan polutan selama jam sibuk atau saat ventilasi kurang optimal.



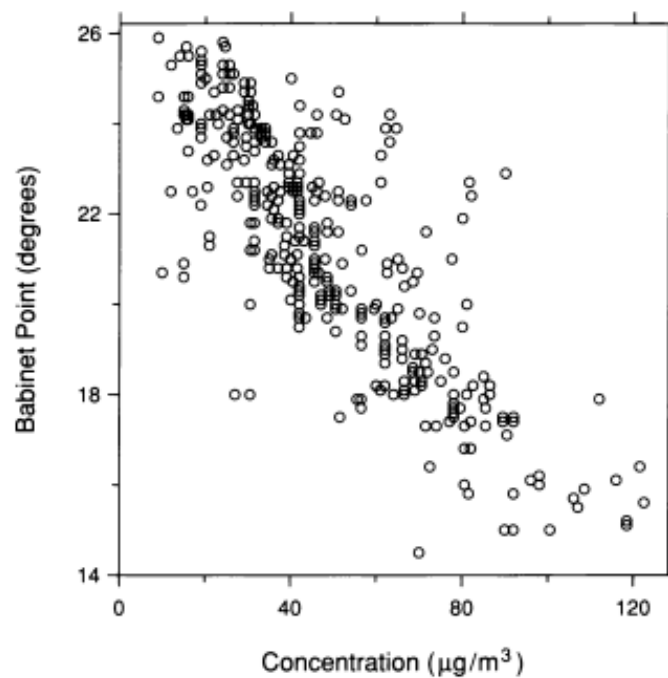
Gambar 7 Contoh Grafik Linear (Sumber: [33])

Grafik linear memberikan gambaran yang lebih jelas dibandingkan tabel data mentah karena manusia lebih mudah mengenali pola dari representasi visual dibandingkan dengan angka-angka dalam tabel. Grafik ini cocok untuk menunjukkan tren perubahan kualitas udara dari waktu ke waktu, sehingga memudahkan identifikasi pola yang signifikan [8]. Selain itu, grafik linear dapat digunakan untuk membandingkan beberapa variabel secara bersamaan, misalnya membandingkan kadar CO₂ di beberapa ruangan berbeda dalam satu gedung untuk memahami area dengan sirkulasi udara yang lebih baik atau lebih buruk. Grafik ini juga sering digunakan dalam evaluasi efektivitas suatu kebijakan, seperti pengaruh peningkatan ventilasi terhadap penurunan kadar polutan dalam jangka waktu tertentu.

C. *Scatter Plot*

Scatter plot adalah alat utama dalam analisis hubungan antar variabel karena memungkinkan identifikasi pola secara visual sebelum dilakukan analisis statistik lebih lanjut [8]. *Scatter plot* menampilkan hubungan antara dua variabel dalam bentuk titik-titik pada bidang kartesian. Teknik ini sangat berguna untuk mengeksplorasi korelasi antara dua parameter dalam data kualitas udara, seperti hubungan antara kelembaban dan kadar CO₂ atau suhu dan PM_{2.5}.

Sebagai contoh, *scatter plot* dapat digunakan untuk melihat apakah kelembaban tinggi berkontribusi terhadap penurunan kadar CO₂ di suatu ruangan. Jika terdapat pola tertentu, seperti titik-titik yang membentuk garis lurus atau kurva, ini dapat menunjukkan adanya korelasi positif atau negatif antara kedua variabel tersebut.



Gambar 8 Contoh *Scatter plot* (Sumber: [8])

Scatter plot juga sering digunakan untuk analisis multivariat, dimana beberapa *scatter plot* dapat dibandingkan untuk mengidentifikasi hubungan kompleks antara tiga atau lebih variabel. Sebagai contoh, dengan menambahkan ukuran atau warna titik berdasarkan variabel tambahan seperti waktu atau lokasi, *scatter plot* dapat memberikan wawasan lebih mendalam tentang interaksi antara berbagai faktor yang mempengaruhi kualitas udara.

2.8 Interpolasi Data

Data hilang (*missing data*) adalah masalah umum dalam analisis data dari berbagai bidang. *Interpolasi* sebagai salah satu teknik imputasi, terutama untuk data deret waktu (*time series*) atau data spasial yang memiliki struktur autokorelasi, dimana nilai pada suatu titik dapat diperkirakan dari nilai-nilai titik di sekitarnya (dalam waktu atau ruang). Prinsip dasar interpolasi adalah mengestimasi nilai yang hilang dengan memanfaatkan informasi dari titik-titik data lain yang tersedia[34].

Metode interpolasi linear merupakan teknik paling sederhana dan paling umum digunakan untuk mengestimasi nilai yang hilang antara dua titik data yang diketahui. Metode ini cocok untuk data yang perubahan antar titiknya relatif stabil dan mengikuti pola linear. Konsep dasarnya adalah menghubungkan dua titik data dengan garis lurus dan mengambil nilai pada titik tengah di antara kedua titik tersebut. Rumus umum interpolasi linear adalah sebagai berikut:

$$f(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{(x_1 - x_0)}(x - x_0) \quad (8)$$

Keterangan:

$f(x)$: Nilai fungsi dari titik yang dicari

$f(x_0)$: Nilai fungsi dari titik awal

$f(x_1)$: Nilai fungsi dari titik ujung

x_0 : Titik Awal

x_1 : Titik Ujung

2.9 Evaluasi Model

Evaluasi model prediksi berbasis *Long Short-Term Memory* (LSTM) dalam data deret waktu (time series) sangat penting untuk mengukur kinerja dan akurasi model dalam memprediksi nilai di masa depan. Beberapa metrik yang umum digunakan dalam evaluasi model LSTM meliputi *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), *Mean Squared Error* (MSE) dan *Root Mean Squared Error* (RMSE). Nilai metrik yang lebih kecil mencerminkan kesalahan prediksi yang lebih rendah dan kinerja model yang lebih baik, sedangkan nilai metrik yang semakin besar menunjukkan penurunan akurasi model.

1. *Mean Absolute Percentage Error* (MAPE)

Mean Absolute Percentage Error (MAPE) adalah metrik yang banyak dipakai dalam evaluasi kinerja model peramalan deret waktu. Konsep utamanya adalah mengukur rata-rata selisih absolut antara nilai ramalan dan nilai aktual, dinyatakan dalam persentase terhadap nilai aktual tersebut. Secara ilmiah, MAPE dirumuskan sebagai berikut:

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (9)$$

Dengan keterangan:

A_t : Nilai aktual pada waktu t ,

F_t : Nilai ramalan pada waktu t

n : Jumlah total pengamatan.

2. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) adalah salah satu metrik evaluasi yang sering digunakan untuk mengukur akurasi prediksi dalam berbagai model regresi, termasuk time series forecasting seperti prediksi kualitas udara dengan menggunakan LSTM (Long Short-Term Memory). MAE memberikan gambaran yang jelas tentang seberapa jauh prediksi model dari nilai yang sebenarnya (aktual), dengan menghitung rata-rata dari kesalahan absolut antara nilai prediksi dan nilai yang sebenarnya. Persamaan untuk menghitung nilai MAE standar ditunjukkan dalam persamaan berikut.

$$MAE = \frac{1}{n} \sum_{i=1}^n P_i - A_i \quad (10)$$

Dengan keterangan:

P_i : Nilai prediksi model untuk sampel ke-iii

A_i : Nilai aktual (atau nilai yang benar) untuk sampel ke-i

n : Jumlah total sampel atau data point yang digunakan dalam perhitungan.

3. Mean Squared Error (MSE)

Mean Squared Error merupakan metrik evaluasi yang mengukur rata-rata kesalahan kuadrat antara nilai yang diprediksi dan nilai aktual. MSE sangat berguna karena memberikan penalti lebih besar terhadap kesalahan yang lebih besar, sehingga model dapat lebih optimal dalam melakukan prediksi. MSE didefinisikan sebagai berikut:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (11)$$

Dengan keterangan:

n : Jumlah sampel data,

y_i : Nilai aktual, dan

$\hat{y}_i$: Nilai prediksi model.

Semakin kecil nilai MSE, semakin baik kinerja model karena menunjukkan bahwa rata-rata kesalahan kuadrat antara prediksi dan nilai aktual lebih kecil. Namun, karena MSE menggunakan skala kuadrat dari satuan data asli, interpretasinya menjadi kurang intuitif.

4. Root Mean Squared Error (RMSE)

Root Mean Squared Error adalah akar kuadrat dari MSE, yang membuat metrik ini memiliki satuan yang sama dengan data asli. RMSE lebih sering digunakan karena lebih mudah diinterpretasikan dibandingkan dengan MSE. RMSE dihitung dengan rumus berikut:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (12)$$

Dengan keterangan:

n : Jumlah sampel data,

y_i : Nilai aktual, dan

$\hat{y}_i$: Nilai prediksi model.

RMSE memberikan bobot lebih besar pada kesalahan yang lebih besar, sehingga lebih sensitif terhadap outlier dibandingkan dengan Mean Absolute Error (MAE). Dalam konteks prediksi kualitas udara, RMSE yang rendah menunjukkan bahwa model mampu memprediksi tren dan nilai kualitas udara dengan lebih akurat.

Tabel 5 Perbandingan Tiap Metrik Evaluasi

Metrik	Keunggulan	Kekurangan
MAPE	Interpretasi langsung, sama satuannya dengan data asli dan tidak terlalu dipengaruhi outlier moderat	Tidak membedakan arah kesalahan serta kesalahan besar dan kecil diberi bobot sama
MSE	Memberi penalti berat pada kesalahan besar serta berguna untuk optimasi model berbasis kuadrat	Kuadrat membuat satuan kurang intuitif dan sangat sensitif terhadap outlier ekstrem
RMSE	Penalti lebih besar pada kesalahan besar serta satuan kembali ke skala asli, lebih intuitif dibanding MSE	Masih sensitif terhadap outlier serta nilai tunggal dapat menutupi distribusi kesalahan yang lebar
MAE	Hasil dalam persentase, mudah dibandingkan antar dataset serta cocok untuk konteks bisnis atau laporan	Tidak terdefinisi jika nilai aktual nol serta bias berlebih saat nilai aktual sangat kecil

Dalam evaluasi model prediksi deret waktu berbasis *Long Short-Term Memory* (LSTM), *Mean Absolute Percentage Error* (MAPE) digunakan sebagai metrik utama karena memberikan interpretasi yang mudah dipahami dalam bentuk persentase kesalahan rata-rata. MAPE memungkinkan penilaian akurasi prediksi secara relatif terhadap nilai aktual, sehingga sangat berguna dalam konteks aplikasi kualitas udara yang memerlukan gambaran tingkat kesalahan yang mudah diinterpretasikan oleh pemangku kepentingan.

Berdasarkan penelitian De Guzman [35], nilai MAPE di bawah 5% dianggap menunjukkan performa model yang sangat baik, sedangkan nilai antara 5% hingga 10% menunjukkan tingkat akurasi yang baik dan dapat diterima dalam banyak aplikasi praktis. Nilai MAPE yang lebih tinggi dari 10% menandakan perlunya perbaikan model atau pengayaan data.

Selain MAPE, metrik lain seperti Mean Absolute Error (MAE), Mean Squared Error (MSE), dan Root Mean Squared Error (RMSE) digunakan sebagai pelengkap untuk memberikan gambaran lebih komprehensif mengenai kesalahan absolut, kesalahan kuadrat, dan sensitivitas model terhadap outlier. MAE menggambarkan kesalahan rata-rata dalam satuan yang sama dengan data asli, sedangkan RMSE memberikan penalti lebih berat pada kesalahan besar sehingga

membantu mendeteksi deviasi ekstrem. MSE, meskipun kurang intuitif karena satuannya adalah kuadrat dari data asli, berguna dalam proses optimasi model.

Penggunaan MAPE sebagai metrik utama dengan dukungan metrik tambahan ini memungkinkan evaluasi yang holistik dan akurat terhadap kinerja model prediksi CO₂ dalam konteks kualitas udara dalam ruang kelas.

2.10 Simpangan Baku (*Standard Deviation*)

Simpangan baku (*standard deviation*) adalah ukuran statistik yang digunakan untuk menggambarkan sebaran data terhadap nilai rata-ratanya. Nilai simpangan baku yang lebih kecil menunjukkan data lebih stabil dan homogen, sedangkan nilai yang lebih besar menunjukkan adanya variasi yang tinggi [36]. Secara matematis, simpangan baku didefinisikan sebagai akar kuadrat dari varians, dengan persamaan berikut:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (13)$$

Keterangan:

σ : Simpangan Baku

x_i : Nilai data ke- i

$\bar{x}$: Rata-rata data

N : Jumlah pengamatan

Dalam bidang rekayasa lingkungan dan studi kenyamanan termal, simpangan baku sering digunakan sebagai indikator stabilitas kondisi fisik, termasuk suhu ruangan. Selain nilai simpangan baku itu sendiri, perbandingan antar kategori ruangan dapat dianalisis dengan menghitung persentase perbaikan stabilitas menggunakan persamaan:

$$\text{Improv}_{\sigma} = \left(1 - \frac{\sigma_{\text{perm}}}{\sigma_{\text{temp}}}\right) \times 100\% \quad (14)$$

Keterangan:

Improv_{σ} : Persentase perbaikan stabilitas suhu

$\frac{\sigma_{\text{perm}}}{\sigma_{\text{temp}}}$: Simpangan baku perbandingan

2.11 Python

Python adalah bahasa pemrograman yang sangat populer, terutama dalam bidang Machine Learning dan Data Science. Bahasa ini pertama kali diperkenalkan oleh Guido van Rossum pada tahun 1991 dan terus berkembang hingga saat ini. Python memiliki sintaks yang sederhana, mudah dibaca, serta mendukung berbagai paradigma pemrograman, baik prosedural, berbasis objek (OOP), maupun fungsional. Python bersifat open-source dan mendukung berbagai sistem operasi, seperti Windows, Linux, dan macOS, yang membuatnya fleksibel untuk berbagai keperluan pengembangan.

Python memiliki pustaka yang sangat lengkap dan ekosistem yang luas. Beberapa pustaka penting dalam bidang ini meliputi NumPy untuk perhitungan numerik, Pandas untuk manipulasi data, Matplotlib dan Seaborn untuk visualisasi data, serta TensorFlow dan PyTorch untuk implementasi model pembelajaran mesin dan *deep learning*[37]. Berkat fleksibilitasnya, Python banyak digunakan dalam berbagai bidang, termasuk pengembangan web (dengan *framework* seperti Django dan Flask), *Data analytics*, pengolahan citra dan suara, kecerdasan buatan, serta pemodelan ilmiah. Dalam penelitian kualitas udara, Python digunakan untuk analisis data eksploratif, pembuatan model prediksi berbasis LSTM, serta visualisasi hasil prediksi untuk mendukung pengambilan keputusan berbasis data.

2.12 Streamlit

Streamlit adalah sebuah *framework* yang memungkinkan pengembang mengubah kode Python langsung menjadi aplikasi web interaktif tanpa perlu menulis kode frontend. Setiap kali pengguna berinteraksi melalui input, klik, slider, atau elemen lainnya, Streamlit akan menjalankan ulang kode dari atas hingga bawah dengan nilai input terbaru, sehingga hasil yang ditampilkan selalu diperbarui secara real-time. Dalam konteks penelitian ini, Streamlit berperan sebagai *dashboard* interaktif yang memungkinkan pengguna untuk memantau dan menganalisis kualitas udara secara real-time. Keunggulan utama Streamlit terletak pada kemudahannya dalam integrasi dengan model machine learning, termasuk LSTM

untuk prediksi kualitas udara. Dengan fitur seperti grafik linier, *scatter plot*, dan *heatmap*, Streamlit mampu menampilkan tren polusi udara serta hubungan antar parameter lingkungan dengan lebih intuitif [38].

2.13 Google Colaboratory

Google Colab adalah sebuah Integrated Development Environment (IDE) untuk pemrograman Python yang menawarkan kemampuan pemrosesan yang dilakukan oleh server Google dengan perangkat keras berperforma tinggi. Selain itu, Google Colab juga merupakan platform berbasis cloud yang memungkinkan pengguna untuk menjalankan kode secara bersamaan. Platform ini memberikan akses gratis ke GPU (Graphics Processing Unit) dan TPU (Tensor Processing Unit) yang sangat penting untuk melatih model pembelajaran mesin [39].

2.14 Penelitian Terkait

Dalam penyusunan skripsi ini terinspirasi dari berbagai penelitian sebelumnya yang membahas teknik *data analytics*, algoritma *Long Short-Term Memory* serta teknik visualisasi data eksploratif. Literatur-literatur tersebut digunakan sebagai acuan untuk memperkuat landasan teori, memperluas pemahaman, dan mengembangkan ide-ide baru yang relevan dengan penelitian ini.

Berdasarkan penelitian terkait yang telah dikaji, berbagai pendekatan dalam prediksi kualitas udara telah dikembangkan menggunakan metode *machine learning* dan *deep learning*. Penelitian oleh Omoniyi Samuel Aiyenuro dan Oluwatomisin Ajayi data kualitas udara dianalisis menggunakan *Exploratory Data Analysis* (EDA) untuk mengidentifikasi pola dan tren dalam data polusi udara di Nigeria. Penelitian ini mengungkapkan bahwa pendekatan EDA memungkinkan pengelolaan data polutan dengan lebih baik dan memfasilitasi pembuatan aplikasi manajemen kualitas udara yang lebih akurat [40].

Penelitian yang dilakukan oleh Afonso Mota, Carlos Serôdio, Ana Briga-Sá, dan Antonio Valente berfokus pada implementasi arsitektur Internet of Things (IoT) untuk memantau kualitas udara dalam ruangan, dengan studi kasus yang dilakukan selama periode tidur. Data yang diperoleh dari sensor kemudian dikirimkan melalui MQTT ke server InfluxDB untuk disimpan dan dianalisis menggunakan Python [41].

Selain itu, model Multi-layer LSTM yang digunakan oleh Yahya Koçak dan Murat Koklu dalam prediksi kualitas udara di Provinsi Konya menunjukkan hasil yang lebih baik dibandingkan LSTM satu lapis, di mana terjadi penurunan nilai RMSE dan MAPE yang signifikan [42]. Studi oleh Lisa Kristiana dan Dian Miyanto menambahkan parameter PM2.5 dalam model LSTM berbasis RNN, membuktikan efektivitas model dalam menangani data sekuensial [43].

Lebih lanjut, Yadi Karyadi dan Handri Santoso membandingkan performa LSTM, Bidirectional LSTM, dan GRU, dengan hasil menunjukkan bahwa model berbasis LSTM lebih unggul dalam memodelkan data deret waktu dibandingkan GRU [44].

Penelitian oleh Shurui Fan menggabungkan teknik data decomposition dan algoritma prediksi Long Short-Term Memory (LSTM). Pendekatan ini sangat berguna dalam memprediksi polusi udara seperti PM2.5 dan CO2 yang sering menunjukkan pola musiman yang kuat serta fluktuasi jangka pendek yang sulit diprediksi dengan model tradisional [45].

Penelitian oleh Jesús Cáceres-Tello berfokus pada analisis dan prediksi polusi PM2.5 di Madrid dengan menggunakan model hibrida antara dan Long Short-Term Memory (LSTM). Penelitian ini memberikan wawasan yang penting untuk pengambilan keputusan berbasis data, terutama dalam upaya mitigasi polusi udara di perkotaan [46]. Yingbo Zhu menerapkan IoT dan LSTM dalam pemantauan CO2 dalam ruangan [47], Penelitian oleh Edra Arkananta Gultom mengembangkan aplikasi peramalan harga saham menggunakan analisis deret

waktu dengan model . Studi ini mengintegrasikan yfinance untuk pengambilan data, untuk peramalan, dan Plotly untuk visualisasi hasilnya [48].

Terakhir, penelitian oleh Zheng Zhang menggabungkan Wavelet Transform, DCCA Correlation Analysis, dan LSTM, yang menunjukkan peningkatan akurasi prediksi dengan kombinasi teknik prapemrosesan data dan model deep learning [49].

Berdasarkan penelitian-penelitian ini, dapat disimpulkan bahwa LSTM dan variannya memiliki keunggulan dalam menangani data sekuensial dan telah digunakan dalam berbagai studi untuk meningkatkan akurasi prediksi kualitas udara, baik dalam lingkungan luar maupun dalam ruangan.

Tabel 6 Penelitian Terkait

No	Peneliti	Objek	Metode	Hasil
1	Omoniyi Samuel Aiyenuro & Oluwatomisin Ajayi (2024) [40]	Kualitas udara di Ogun State, Nigeria	Eksploratory Data Analysis (EDA), Pengembangan Aplikasi Manajemen Kualitas Udara	Ditemukan korelasi positif antara PM2.5 dan PM10 (0.69), serta antara NO dan NO _x (0.97). Ditemukan korelasi negatif antara kelembaban relatif (RH) dan PM2.5 (-0.46). Pengembangan aplikasi untuk visualisasi data kualitas udara dan analisis sentimen publik
2	Afonso Mota, Carlos Serôdio, Ana Briga-Sá, Antonio	Kualitas udara dalam ruangan selama periode tidur	Arsitektur IoT berbasis ESP32-C6, MQTT,	Tidur dengan pintu sedikit terbuka menyebabkan peningkatan CO ₂ yang

No	Peneliti	Objek	Metode	Hasil
	Valente (2025) [41]		InfluxDB, dan Python	lebih lambat dan tetap di bawah ambang batas yang disarankan, dibandingkan dengan pintu tertutup yang menyebabkan peningkatan CO2 lebih cepat dan melebihi batas aman
3	Yahya Koçak & Murat Koklu (2022) [42]	Prediksi kualitas udara di Provinsi Konya	Multi-layer LSTM	Model multi-layer LSTM menghasilkan prediksi kualitas udara yang lebih akurat dibanding single-layer LSTM. Pada stasiun Karatay, RMSE turun dari 15.14 menjadi 11.61 dan MAPE dari 19.16% menjadi 12.65%.
4	Lisa Kristiana & Dian Miyanto (2023) [43]	Prediksi kualitas udara di Kota Bandung	LSTM berbasis RNN	LSTM mampu memprediksi kualitas udara dengan RMSE sebesar 4.0279.
5	Yadi Karyadi & Handri Santoso (2022) [44]	Perbandingan model prediksi kualitas udara	LSTM, Bidirectional LSTM, & GRU	LSTM dan Bidirectional LSTM lebih akurat dibandingkan GRU dalam memprediksi kualitas udara.
6	Shurui Fan, Dongxia Hao, Yu Feng, Kewen Xia &	Kualitas udara (terutama PM2.5 dan CO2)	Data Decomposition + LSTM	Model hibrida yang menggabungkan teknik dekomposisi data dan LSTM berhasil

No	Peneliti	Objek	Metode	Hasil
	WenbiaoYang (2021) [45]			meningkatkan akurasi prediksi kualitas udara dengan mengurangi noise dan memfokuskan model pada pola yang relevan. Model ini lebih akurat dibandingkan dengan model tradisional dalam memprediksi polusi udara.
7	Jesús Cáceres-Tello & José Javier Galán-Hernández (2024) [46]	Polusi PM2.5 di Madrid	Prophet + LSTM	Model hibrida Prophet - LSTM meningkatkan akurasi prediksi kualitas udara, mampu menangkap pola musiman dan tren jangka panjang (dari Prophet) serta fluktuasi jangka pendek (dari LSTM). Model ini lebih akurat dibandingkan dengan model tradisional dan LSTM tunggal dalam prediksi PM2.5.
8	Yingbo Zhu, Shahriar Abdullah Al-Ahmed, Muhammad Zeeshan Shakir, Joanna Isabelle Olszewska	Pemantauan CO2 dalam ruangan berbasis IoT	IoT & LSTM	Model LSTM mampu memprediksi kadar CO2 dengan margin kesalahan 5.5%.

No	Peneliti	Objek	Metode	Hasil
	(2023) [47]			
9	Edra Arkananta Gultom, Kartika Dewi Sri Susilowati, Anik Kusmintarti (2024) [48]	Harga saham Indonesia	Model Prophet, Streamlit, yfinance, Plotly	Aplikasi peramalan harga saham yang memungkinkan pengguna memilih saham dan menyesuaikan periode prediksi. Model memberikan prediksi jangka pendek yang berguna namun akurasinya terbatas oleh volatilitas pasar dan faktor eksternal
10	Zheng Zhang, Haibo Chen, Xiaoli Huang (2023) [49]	Prediksi kualitas udara dengan kombinasi teknik	Wavelet Transform, DCCA, & LSTM	Kombinasi teknik ini meningkatkan akurasi prediksi sebesar 1-5% dibandingkan LSTM standar.

Penelitian akan berfokus pada penerapan *data analytics* untuk prediksi dan visualisasi kualitas udara dalam ruangan dengan menggunakan algoritma dan *Long Short-Term Memory* (LSTM) guna mendukung pengambilan keputusan berbasis data. Dengan meningkatnya kesadaran akan pentingnya kualitas udara dalam ruangan, penelitian ini bertujuan untuk mengembangkan model prediksi yang mampu mengidentifikasi pola perubahan polutan udara, seperti CO₂, PM_{2.5}, PM₁₀, suhu, dan kelembaban, berdasarkan data historis yang tersedia. Selain itu,

penelitian ini juga menekankan pentingnya visualisasi data eksploratif, yang dilakukan melalui *dashboard* semi interaktif berbasis Streamlit, guna memberikan wawasan yang lebih intuitif kepada pemangku kepentingan terkait kondisi kualitas udara dalam ruangan.

Dalam proses analisis data, penelitian ini menerapkan metode OSEMN (Obtain, Scrub, Explore, Model, iNterpret) untuk mengelola data secara sistematis, mulai dari pengumpulan dan pembersihan data, eksplorasi untuk menemukan pola tren, pembangunan model LSTM, hingga interpretasi hasil prediksi. Evaluasi model dilakukan menggunakan metrik seperti Mean Absolute Error (MAE), Mean Squared Error (MSE), dan Root Mean Squared Error (RMSE) untuk memastikan akurasi prediksi yang dihasilkan. Hasil penelitian diharapkan dapat memberikan kontribusi dalam pengelolaan kualitas udara, baik dalam konteks akademik maupun implementasi praktis, dengan memungkinkan pengguna untuk mengambil langkah-langkah preventif dalam menjaga kualitas udara dalam ruangan melalui prediksi berbasis data historis.

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat

Penelitian dan pembuatan skripsi ini dilakukan pada:

Waktu : Maret 2025 – Juli 2025

Tempat : Laboratorium Komputer, Teknik Elektro
Universitas Lampung

Tabel 7 Aktivitas Penelitian

No.	Aktivitas	Maret				April				Mei				Juni				Juli			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1.	Studi Literatur																				
2.	Obtain																				
3.	Scrub																				
4.	Explore																				
5.	Modelling																				
6.	Evaluasi																				
7.	Interpret																				
8.	Analisis Pembahasan																				

3.2 Alat dan Bahan

Terdapat alat dan bahan yang dibutuhkan dalam penelitian ini guna membantu penulis dalam penyusunan tugas akhir yang dapat diuraikan berikut ini:

3.2.1 Alat Penelitian

Penelitian ini menggunakan perangkat dengan spesifikasi berikut:

1. Perangkat keras (*Hardware*)

Perangkat keras yang digunakan selama penelitian yaitu laptop dengan spesifikasi sebagai berikut:

Tabel 8 *Hardware*

No.	Kebutuhan	Spesifikasi	Kegunaan
1.	Laptop	Intel(R) Core(TM) i5- 1035G1 CPU @ 1.000 GHz (8CPUs) RAM 8 GB	Laptop ini berfungsi sebagai perangkat utama untuk menjalankan proses komputasi dan pengolahan data yang mendukung operasional sistem serta pelatihan model Machine Learning. Selain itu, perangkat ini digunakan untuk analisis data secara komprehensif dan penyusunan laporan penelitian dengan metode yang sistematis dan terstruktur.

2. Perangkat Lunak (*Software*)

Perangkat lunak yang digunakan selama penelitian sebagai berikut:

Tabel 9 *Software*

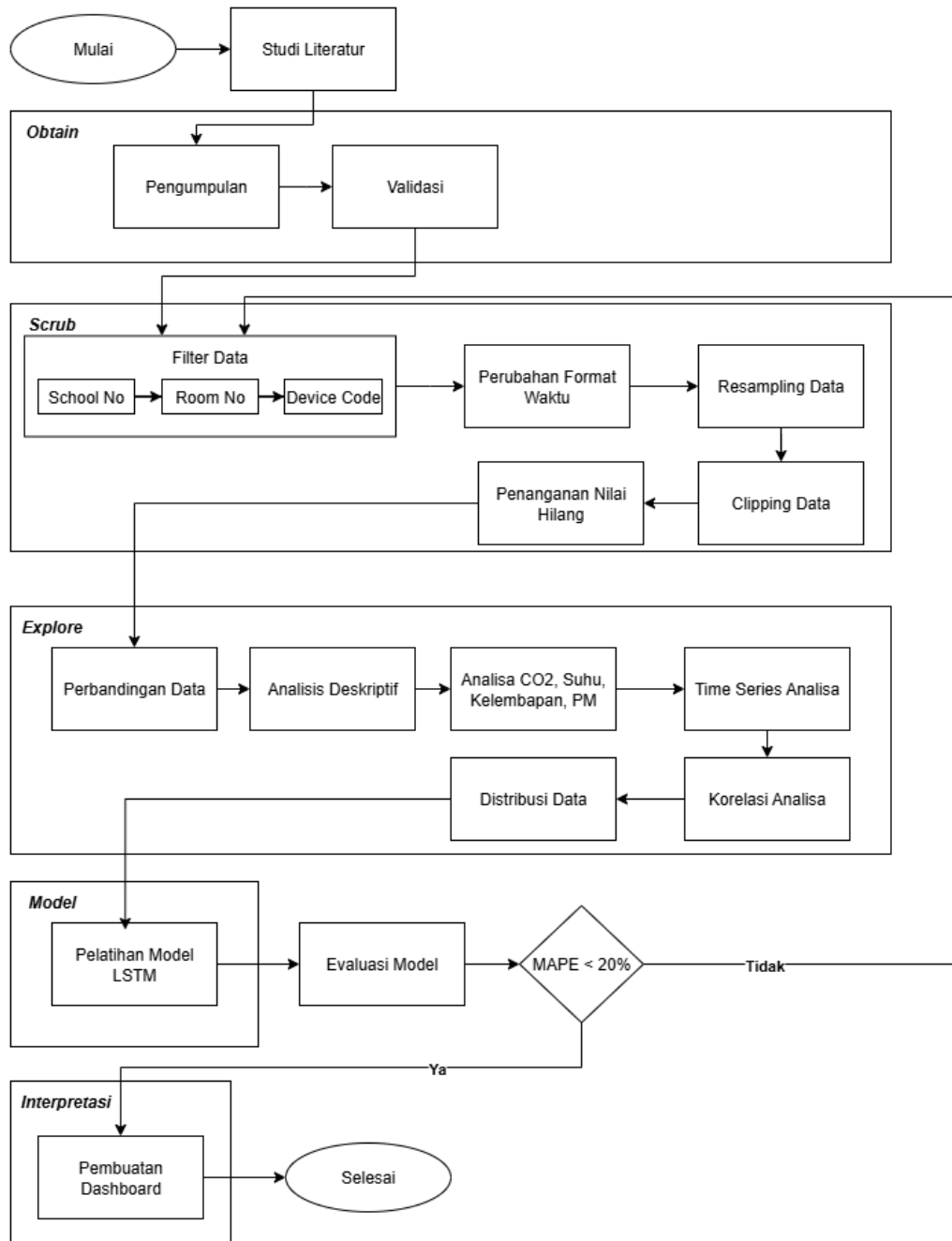
No.	Kebutuhan	Versi	Kegunaan
1.	Python	3.11.11	Bahasa pemrograman utama untuk pengolahan data, pembuatan model LSTM, dan visualisasi hasil prediksi.
	a. Matplotlib	-	Library untuk visualisasi data dalam bentuk grafik dan plot.
	b. Seaborn	-	Library visualisasi statistik berbasis Matplotlib dengan tampilan yang lebih menarik.
	c. Tensorflow	-	<i>Framework</i> untuk pengembangan dan pelatihan model machine learning, khususnya deep learning
	d. Skicitlearn	-	Library machine learning untuk preprocessing, evaluasi, dan validasi model.

No.	Kebutuhan	Versi	Kegunaan
	e. Numpy	-	Library komputasi numerik dan manipulasi array multidimensi.
	f. Pandas	-	Library untuk manipulasi dan analisis data berbasis struktur DataFrame
	g. Streamlit	-	<i>Framework</i> Python untuk membangun aplikasi web interaktif dengan cepat, terutama untuk visualisasi data dan <i>dashboard</i> analitik.
	h. Plotly	-	Library visualisasi interaktif berbasis web yang mendukung pembuatan grafik dinamis seperti line chart, bar chart, <i>scatter plot</i> , dan lainnya dengan kemampuan zoom dan hover.
2.	Google Colaboratory	-	Platform berbasis cloud untuk menjalankan kode Python, melatih model Machine Learning, dan menganalisis data tanpa perlu instalasi lokal
3.	<i>Operating System</i>	Windows 11 64-bit	Menyediakan lingkungan perangkat lunak untuk menjalankan aplikasi, termasuk Python dan pustaka pendukung untuk pemodelan LSTM.

3.3 Tahapan Penelitian

Penelitian ini diawali dengan studi literatur untuk memperoleh pemahaman mendalam mengenai konsep, teori, serta metode yang relevan. Selanjutnya, penelitian mengikuti tahapan dalam metode OSEMN (*Obtain, Scrub, Explore, Model, dan Interpret*). Pada tahap *Obtain*, dilakukan pengumpulan data historis kualitas udara dalam ruangan dari sumber yang telah ditentukan. Dilanjutkan tahap *Scrub* melibatkan pembersihan data dan transformasi agar dataset siap digunakan dalam pengembangan model prediktif. Pada tahap *Explore* dilakukan untuk analisis eksploratif guna memahami pola dan karakteristik data. Dilanjutkan ke tahap *Model*, diterapkan metode LSTM untuk membangun model prediksi kualitas udara berdasarkan pola data historis. Model yang dihasilkan kemudian dievaluasi pada tahap *Evaluasi*, dimana performanya diukur menggunakan metrik evaluasi utama yaitu *Mean Absolute Percentage Error* (MAPE).

Setelah model mencapai tingkat akurasi yang diinginkan, tahap akhir adalah interpret, dimana hasil prediksi divisualisasikan untuk mendukung pengambilan keputusan. Visualisasi ini bertujuan untuk menyajikan informasi yang mudah diinterpretasikan oleh pengguna, sehingga dapat membantu dalam perencanaan dan pengelolaan kualitas udara dalam ruangan secara lebih efektif. Alur kerja penelitian ini dapat dilihat pada Gambar 9.



Gambar 9. Flowchart OSEM

3.3.1. Obtain

Tahapan Obtain dalam penelitian ini mencakup proses pengumpulan data yang digunakan untuk analisis dan prediksi kualitas udara dalam ruangan. Dataset yang digunakan dalam penelitian ini berasal dari penelitian yang diteliti oleh Rita Elise van der Walt [50]. Dataset ini menyediakan data hasil pemantauan kualitas udara dalam berbagai ruang kelas di sekolah Afrika Selatan dengan karakteristik bangunan yang berbeda. Dataset ini terdiri dari 19 kolom data yang mencakup berbagai parameter lingkungan. Penyebaran kolom dapat dilihat pada Tabel 10.

Tabel 10 Struktur Dataset

No	Kolom data	Satuan	Deskripsi
1	Date	YYYY-MM-DD	Waktu pengambilan data untuk setiap parameter.
2	Time	HH:MM:SS	Waktu spesifik dalam format jam-menit-detik.
3	Classroom Type	-	Tipe ruangan, seperti kontainer atau ruang dengan isolasi.
4	Classroom Category	-	Kategori kelas berdasarkan jenis ruangan (misalnya permanen atau sementara).
5	Room Number	-	Nomor ruangan yang menjadi lokasi pengambilan data.
6	Device Code	-	Kode perangkat sensor yang digunakan untuk pengukuran.
7	School No	-	Nomor sekolah yang terkait.
8	Grade	-	Kelas (0 hingga 7, dengan 0 untuk pendidikan pra-sekolah).
9	Measured T	°C	Suhu udara dalam ruangan yang mempengaruhi kenyamanan termal.
10	Measured RH	%	Tingkat kelembapan udara dalam ruangan.
11	Measured CO2	ppm	Kandungan karbon dioksida dalam ruangan, yang berpengaruh terhadap kenyamanan dan kesehatan.
12	Measured PM1.0	µg/m ³	Konsentrasi partikel halus dengan ukuran ≤ 1 mikrometer yang dapat mempengaruhi kesehatan pernapasan.
13	Measured PM2.5	µg/m ³	Konsentrasi partikel kasar dengan ukuran ≤ 2.5 mikrometer yang dapat menyebabkan gangguan kesehatan.
14	Measured	µg/m ³	Konsentrasi partikel kasar dengan

No	Kolom data	Satuan	Deskripsi
	PM10		ukuran ≤ 10 mikrometer yang dapat menyebabkan gangguan kesehatan.
15	School Day	-	Indikasi apakah hari tersebut adalah hari sekolah.
16	School Hours	-	Durasi jam sekolah.
17	Break Time	-	Waktu istirahat di hari sekolah.
18	Power On	-	Status daya perangkat (apakah perangkat dalam keadaan menyala atau mati).
19	Outdoor Temperature	°C	Suhu luar ruangan yang diukur dalam derajat Celsius.

Data tersebut dikumpulkan dalam total 372.084 baris data selama periode sepuluh bulan, dimulai pada tanggal 8 Februari 2023 hingga 7 September 2023. Pengambilan data tidak dilakukan dengan distribusi yang seragam, melainkan sesuai dengan sensor dalam menerima data pada setiap periode waktu. Sebagai contoh, pada tanggal 8 Februari 2023, diperoleh 71 baris data dengan rentang temporal mulai pukul 07.00 hingga 11.59 sedangkan di keesokan hari nya hanya memperoleh data sebanyak 65 baris data. Pengambilan data dilakukan melalui sensor yang ditempatkan di berbagai ruangan dalam dua lokasi penelitian seperti terlihat pada Gambar 10.



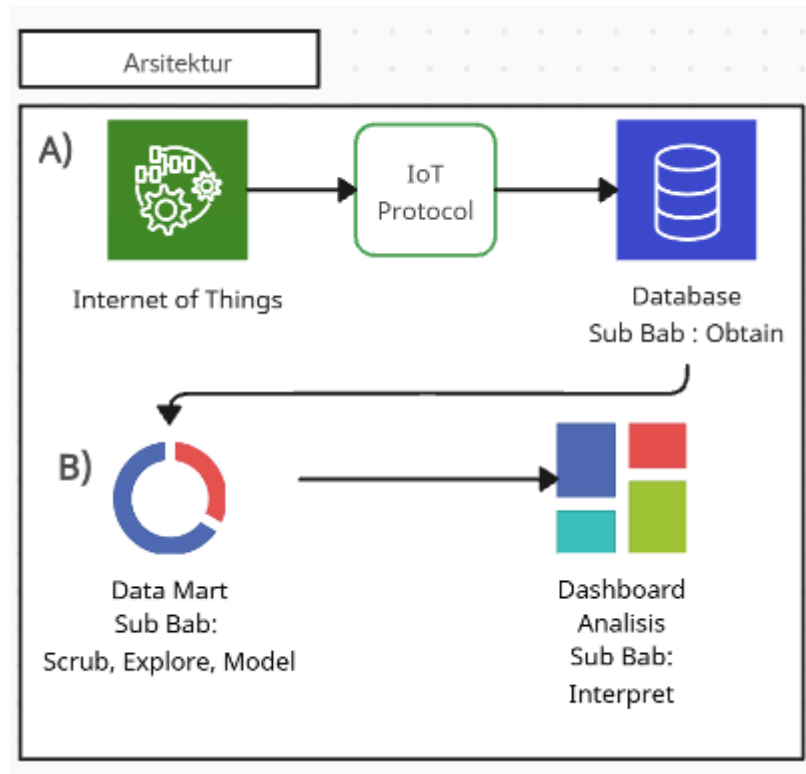
(a)



(b)

Gambar 10 Lokasi ruang kelas dan infrastruktur sekolah (Sumber : [50])

(a) Sekolah-1 (b) Sekolah-2



Gambar 11 Arsitektur Pengerjaan

(A) Keseluruhan Arsitektur (B) Penelitian yang dilakukan

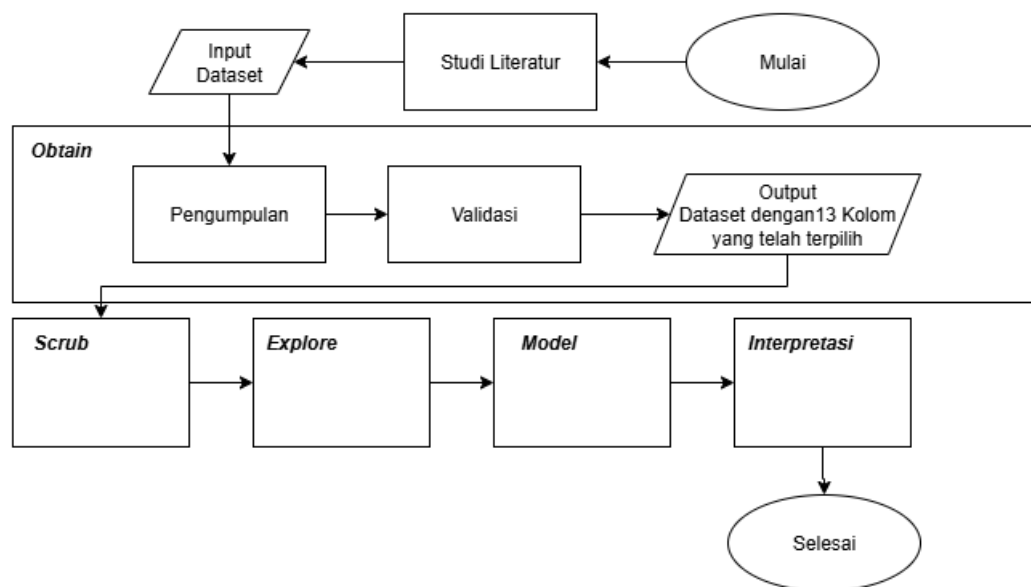
Pengambilan data diawali menggunakan sensor dari setiap ruangan kemudian dikirimkan menggunakan IOT protokol dilanjutkan dimasukan ke dalam database seperti pada Gambar 11. Proses pengambilan data ini disimpan di dalam database sehingga dapat dianalisis.

Dalam penelitian ini, dataset ini digunakan sebagai sumber utama untuk menganalisis pola kualitas udara dalam ruangan. Dataset tersebut diambil dari database kemudian divalidasi apakah dapat dilanjutkan untuk dapat diolah lebih lanjut melalui tahap *Scrub* atau tidak berdasarkan kriteria data yang baik. Batasan sistem berada pada komponen *Data Mart* dan *Dashboard Analisis* seperti yang dapat dilihat pada Gambar 11 (B), di mana penulis berfokus pada pengolahan dan penyajian data kualitas udara yang diterima dari perangkat sensor IoT. Data dari database disaring, dibersihkan, dan diubah menjadi informasi bermakna yang ditampilkan dalam bentuk visual interaktif pada *dashboard*. *Dashboard* ini bertujuan membantu pihak pengelola gedung atau sekolah untuk memantau dan

mengambil keputusan terkait kualitas udara dalam ruangan secara cepat dan informatif. Sebelum melakukan pembuatan *dashboard* diperlukan untuk memilah kolom yang akan digunakan dan dianalisis.

Setelah pengambilan data, tahap selanjutnya adalah validasi data. Pada tahap ini, data diperiksa dan dianalisis untuk menentukan kolom atau parameter mana yang relevan dan dapat digunakan dalam proses analisis lebih lanjut. Validasi ini penting untuk memastikan bahwa hanya data yang sesuai dan konsisten yang digunakan, sehingga hasil analisis dapat diandalkan dan memberikan wawasan yang akurat mengenai kualitas udara serta faktor-faktor yang mempengaruhinya.

Dengan data yang kaya dan bervariasi, penelitian ini bertujuan untuk memberikan wawasan mengenai faktor-faktor yang memengaruhi kualitas udara serta mendukung pengambilan keputusan berbasis data untuk menciptakan lingkungan dalam ruangan yang lebih sehat.



Gambar 12 Hasil Proses Obtain

Setelah tahap Obtain dilakukan, proses ini menghasilkan dataset yang telah tervalidasi, yang mencakup kolom-kolom yang relevan untuk diproses lebih lanjut seperti yang dapat dilihat pada Gambar 12.

Pada tahap Obtain, kolom-kolom yang relevan dengan analisis diidentifikasi dan diseleksi. Kolom-kolom terpilih ini kemudian diteruskan ke tahap selanjutnya dalam pipeline, yaitu proses *scrub*, untuk dibersihkan dan dipersiapkan agar dapat digunakan dalam analisis yang lebih mendalam.

Dengan demikian, output dari tahap Obtain adalah dataset yang telah tervalidasi dan terstruktur dengan baik, siap untuk diproses lebih lanjut, sehingga hasil analisis yang diperoleh lebih akurat dan dapat dipercaya.

3.3.2. *Scrub*

Tahap Scrub dilakukan untuk memastikan bahwa dataset yang digunakan memiliki format yang sesuai dan terbebas dari inkonsistensi data yang dapat mempengaruhi hasil analisis. Proses utama dalam tahap ini meliputi penyesuaian format waktu, filter data, *resampling* data, penanganan *missing values*, dan *clipping* data.

Langkah pertama dalam proses ini adalah memfilter data dimana data dipisahkan berdasarkan tiga kriteria utama. Pertama, data disaring berdasarkan sekolah, untuk memastikan bahwa hanya data yang relevan dari sekolah yang terkait. Setelah data dipisah berdasarkan sekolah, data dipisahkan kembali berdasarkan nomor ruangan dan kode perangkat (*device code*) agar analisis dapat dilakukan pada data yang spesifik untuk setiap ruangan dengan *device code* yang diteliti. Dengan cara ini, data tetap utuh, namun dapat dipilah sesuai dengan kategori yang relevan untuk analisis lebih lanjut.

Tahap selanjutnya adalah penyesuaian format yaitu menggabungkan kolom waktu dan tanggal menjadi satu kolom timestamp. Hal ini dilakukan untuk mempermudah analisis berbasis *time series* dengan format waktu yang seragam. Dengan mengonversi data waktu ke dalam format *datetime*, model dan analisis dapat lebih mudah mengenali pola temporal dalam data, sehingga prediksi kualitas udara dalam ruangan dapat dilakukan secara lebih akurat.

Tahapan berikutnya adalah *resampling data*, yang bertujuan untuk menyamakan frekuensi waktu agar konsisten dan seragam. Dataset yang memiliki frekuensi waktu yang tidak konsisten diubah menjadi interval waktu yang seragam.

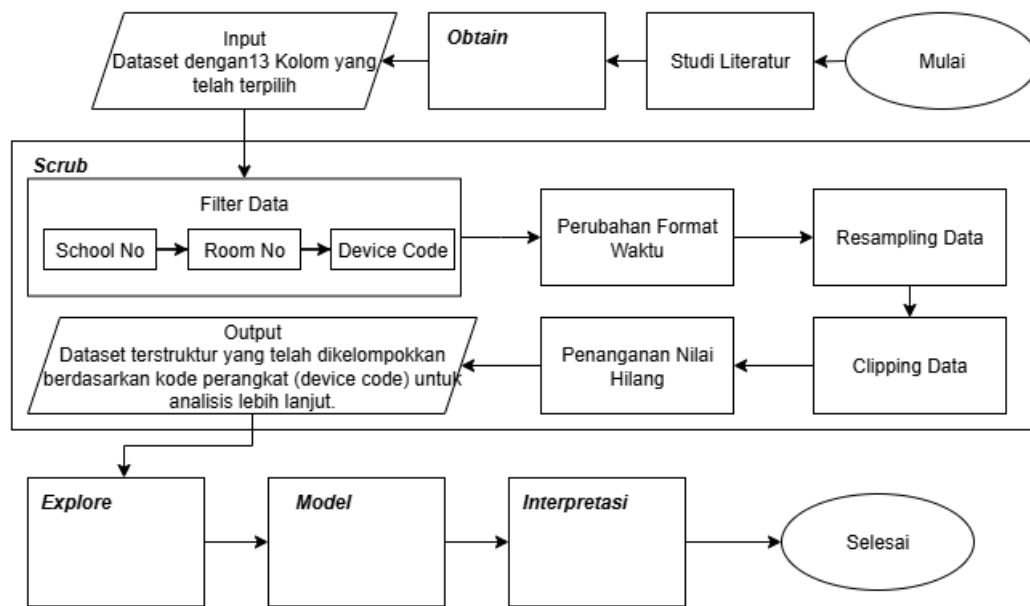
Pemeriksaan terhadap *missing values* dalam setiap kolom data seperti pada Tabel 11. Data kosong dalam dataset dapat mengganggu akurasi model, sehingga perlu dilakukan penanganan yang tepat.

Tabel 11 Kolom data yang berisi Null

Kolom data	Jumlah null	Persen
Measured T	5	0.0013%
Measured RH	2	0.0005%
Measured CO2	33889	9.1%
Measured PM1.0	17882	4.8%
Measured PM2.5	17882	4.8%
Measured PM10	17882	4.8%
Power On	1504	0.4042%

Untuk mengatasi hal ini, *missing values* di setiap kolom diisi dengan interpolasi berdasarkan gap. Metode imputasi ini dipilih karena mampu mempertahankan distribusi data tanpa menghilangkan informasi penting yang terkandung dalam dataset. Dengan menggunakan nilai interpolasi sebagai pengganti data yang hilang, dataset tetap konsisten dan dapat digunakan secara optimal dalam pelatihan model.

Clipping data dilakukan untuk mengeliminasi nilai yang berada di luar rentang yang diharapkan. Data ekstrem atau outlier yang dapat mempengaruhi integritas analisis dipangkas. Hal ini penting untuk memastikan bahwa model yang dihasilkan hanya menggunakan data yang relevan dan sesuai dengan kondisi yang diinginkan dalam penelitian ini.



Gambar 13 Hasil dari proses *scrub*

Setelah data dipisah dan dibersihkan, langkah selanjutnya adalah memisahkan dataset mana yang memenuhi kriteria kualitas untuk dilanjutkan ke tahap analisis lebih lanjut. Dataset yang terpilih adalah data yang sudah bersih, terstruktur dengan interval waktu yang konsisten, dan siap untuk dianalisis lebih mendalam guna menghasilkan temuan yang lebih akurat yang ditunjukkan pada Gambar 13.

3.3.3. *Explore*

Pada tahap *Explore*, dilakukan perbandingan data serta analisis eksploratif terhadap dataset untuk memahami pola tren pada data dan hubungan antar variabel. Proses ini mencakup berbagai proses analisa untuk menggali wawasan awal mengenai karakteristik data.

Data yang telah melalui proses *scrub* dibandingkan dengan data sebelum proses *scrub*. Perbandingan ini bertujuan untuk memberikan wawasan mengenai perubahan yang terjadi setelah data dibersihkan, serta untuk mengidentifikasi perbedaan signifikan antara data awal dan data yang telah diproses.

Analisis deskriptif dilakukan setelah perbandingan kedua data, analisis deskriptif dengan menghitung ukuran statistik seperti mean (rata-rata), modus, standar

deviasi, minimum, maksimum, dan kuartil untuk setiap variabel numerik seperti PM2.5, PM10, CO2, suhu, kelembaban, tekanan udara seperti pada Tabel 12.

Tabel 12 Contoh analisis deskriptif

	Kolom CO2
Count	372.084
Mean	550.235
Std	287.620
Min	200
25%	415
50%	459
75%	550
Max	13.364

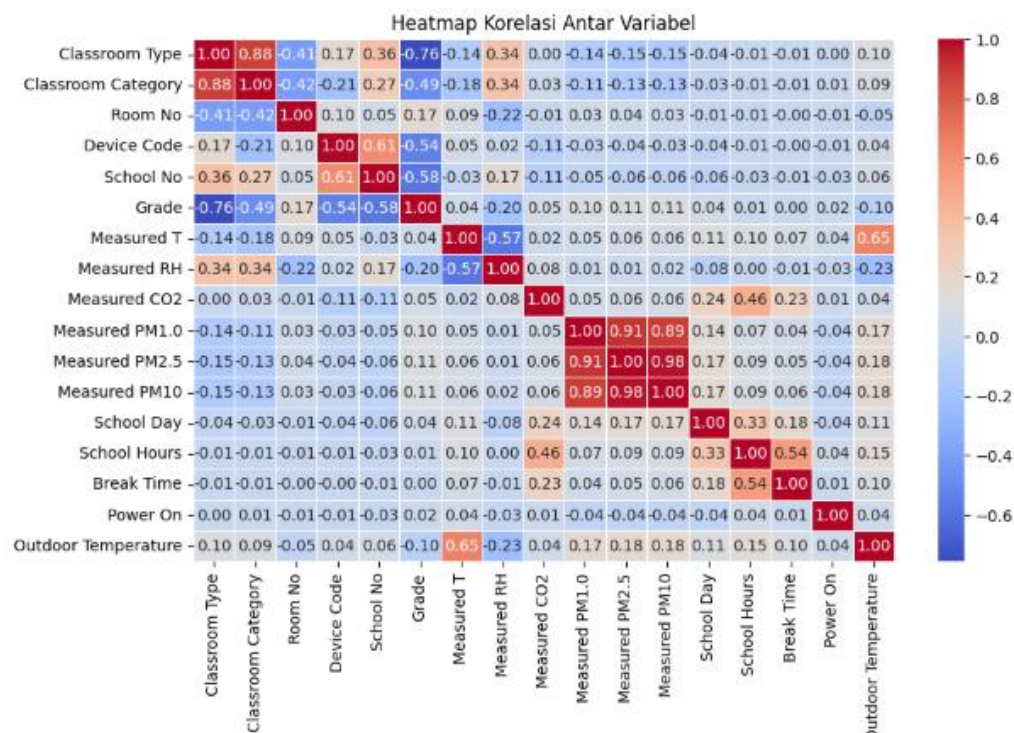
Statistik ini membantu dalam memahami sebaran data serta mendeteksi kemungkinan adanya outlier yang dapat memengaruhi hasil analisis dan menggambarkan rentang dan kecenderungan pusat data sebelum visualisasi.

Tahap berikutnya adalah melakukan visualisasi perbandingan suhu ruangan berdasarkan jenis untuk kedua sekolah yang dianalisis. Visualisasi ini membantu untuk membandingkan pola suhu yang ada di masing-masing sekolah, memberikan gambaran lebih jelas tentang perbedaan suhu antar ruang dengan jenis yang berbeda, serta memungkinkan identifikasi pola tren pada data.

Setelah memperoleh pemahaman terkait pola suhu, langkah selanjutnya adalah melakukan visualisasi untuk konsentrasi CO2 berdasarkan tipe kelas. Visualisasi ini bertujuan untuk menganalisis sebaran konsentrasi CO2 pada masing-masing tipe ruang kelas pada setiap sekolah, sehingga dapat diketahui apakah terdapat perbedaan yang signifikan dalam kualitas udara antar ruang kelas yang berbeda. Dengan melakukan visualisasi terhadap data CO2 maka dapat mengidentifikasi pola faktor-faktor yang memengaruhi akumulasi CO2 di dalam ruang kelas. Pemahaman ini memberikan dasar yang lebih kuat dalam merancang strategi untuk meningkatkan kualitas udara dan menciptakan lingkungan belajar yang lebih sehat dan nyaman.

Pola harian kolom perlu dilakukan eksplorasi sehingga dapat mengetahui jam-jam dimana nilai kadar dari parameter CO₂ dan *particulate matter* meningkat dan melihat perbedaan antara Sekolah-1 dan Sekolah-2. Untuk melihat penyebaran nilai CO₂ kemudian digunakan *scatter plot* 3 dimensi agar dapat mengetahui nilai rata-rata, maksimum, dan minimum CO₂ harian secara intuitif

Selain proses analisa diatas, analisis korelasi perlu dilakukan untuk mengidentifikasi hubungan antara variabel-variabel dalam dataset. Korelasi ini dihitung menggunakan koefisien korelasi Pearson untuk variabel numerik, yang menunjukkan sejauh mana hubungan antara dua variabel bersifat positif atau negatif seperti pada Gambar 14.



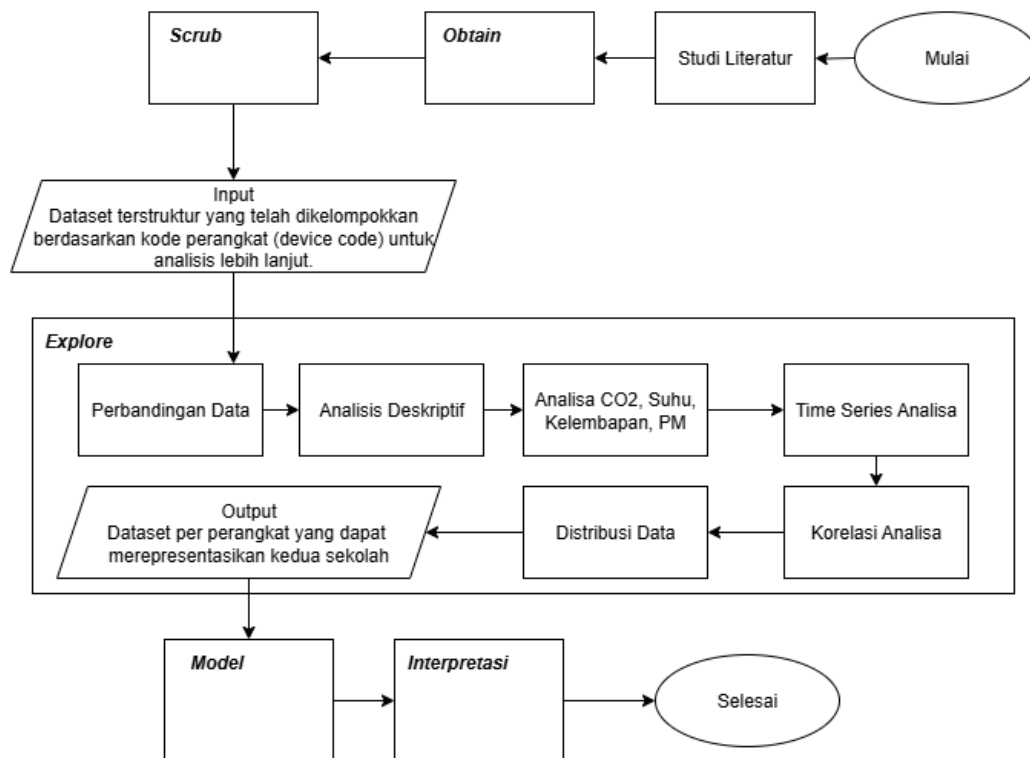
Gambar 14. Korelasi Setiap Kolom

Hubungan positif menonjol yang terlihat pada ketiga parameter partikel (PM1.0, PM2.5, PM10) sangat berkorelasi ($r \geq 0,90$) karena rentang ukurannya saling tumpang-tindih; suhu luar dan suhu dalam ruangan juga berkorelasi moderat ($r = 0,65$), mencerminkan pengaruh cuaca terhadap termal ruang kelas; begitu pula *Classroom Type* dan *Classroom Category* ($r = 0,88$) yang memang dikelompokkan berdasarkan fungsi ruang. Sementara itu, korelasi negatif paling

kuat terlihat antara *Grade* dan *Classroom Type* ($r = -0,76$), menandakan tiap jenjang kelas menempati tipe ruang yang berbeda. Sebagian besar pasangan variabel lainnya memiliki korelasi mendekati nol ($|r| < 0,20$), artinya mereka relatif independen satu sama lain.

Hasil analisis korelasi ini menjadi dasar dalam pemilihan fitur yang paling berpengaruh dalam proses pemodelan prediktif kualitas udara dalam ruangan.

Selain itu, distribusi data diperiksa lebih lanjut untuk memberikan gambaran visual yang lebih jelas mengenai kecenderungan pusat dan rentang data. Distribusi data mengacu pada bagaimana nilai-nilai data tersebar atau terdistribusi dalam rentang yang ada. Dalam analisis eksploratif ini, distribusi data diperiksa untuk memastikan apakah data mengikuti pola tertentu, seperti distribusi normal. Pemeriksaan distribusi dapat dilakukan dengan menggunakan grafik seperti histogram, *boxplot*, atau distribusi kumulatif yang memberikan gambaran visual tentang persebaran data. Serta visualisasi distribusi waktu pengambilan data hingga mendapatkan dataset yang dapat merepresentasikan kedua sekolah untuk melanjutkan pada tahap modeling.

Gambar 15 Hasil Proses *Explore*

Dataset per perangkat diidentifikasi untuk menemukan dataset per perangkat yang dapat merepresentasikan sekolah-1 dan sekolah-2 berdasarkan standar kualitas data. Pemilihan ini didasari bahwa data dari perangkat-perangkat tersebut mampu mencerminkan dinamika kualitas udara secara keseluruhan di sekolah masing-masing.

3.3.4. Modelling

Setelah melalui tahap *explore*, proses penelitian ini berlanjut ke tahap pelatihan model, di mana algoritma *Long Short-Term Memory* (LSTM) digunakan untuk melakukan prediksi kualitas udara dalam ruangan, khususnya untuk memprediksi nilai CO2 berdasarkan data historis. LSTM merupakan arsitektur jaringan saraf tiruan berbasis *Recurrent Neural Network* (RNN) yang dirancang untuk menangani data time series dengan ketergantungan jangka panjang. Model ini

memiliki keunggulan dalam mengatasi masalah *vanishing gradient*, yang sering terjadi pada RNN standar saat menangani data time series yang panjang.

Pemilihan algoritma LSTM dalam penelitian ini didasarkan pada kemampuannya dalam mengenali pola jangka panjang dari data historis kualitas udara. LSTM mampu mengingat dan memproses informasi dalam rangkaian waktu yang lebih panjang dibandingkan metode regresi klasik seperti ARIMA, maupun metode Machine Learning lainnya seperti *Support Vector Regression* (SVR) dan *Random Forest Regression*. Dalam penelitian ini, LSTM digunakan untuk memprediksi variabel kualitas udara, khususnya konsentrasi CO₂, yang dapat membantu dalam memahami tren perubahan kualitas udara di dalam ruangan serta mendukung pengambilan keputusan berbasis data.

Sebelum model LSTM diterapkan, dataset akan dibagi menjadi tiga subset untuk memastikan evaluasi yang optimal. Proses split data dilakukan dengan pembagian 70% untuk data pelatihan, 15% untuk data validasi, dan 15% untuk data uji. Pembagian ini memastikan bahwa model dapat dilatih pada sebagian besar data, dievaluasi pada data yang tidak terlihat sebelumnya selama pelatihan, dan diuji pada data yang sepenuhnya terpisah untuk mengukur performa akhir model.

Arsitektur LSTM terdiri dari lapisan-lapisan jaringan yang dirancang untuk memproses data time series dengan mengingat informasi dalam jangka panjang. Setiap lapisan LSTM memiliki sel memori yang memungkinkan model untuk menyimpan informasi penting dari input sebelumnya dan mempertahankan hubungan antar waktu dalam data. Sel memori ini dikendalikan oleh tiga gerbang utama: *forget gate*, *input gate*, dan *output gate*, yang secara dinamis mengatur informasi mana yang dilupakan, disimpan, atau dikeluarkan. Dengan kemampuan ini, LSTM dapat mengatasi masalah *vanishing gradient* yang biasa terjadi pada RNN konvensional dan mampu menangani ketergantungan jangka panjang dalam data time series.

Model LSTM dalam penelitian ini dioptimasi menggunakan algoritma Adam, yang merupakan salah satu algoritma optimisasi yang efektif untuk jaringan saraf dalam menangani konvergensi cepat dengan pengaturan learning rate yang dinamis. Pengaturan batch size yang efisien juga dipilih untuk mendukung pemrosesan dengan memori terbatas, serta jumlah epoch yang cukup untuk memastikan model belajar secara optimal tanpa overfitting.

3.3.5. Evaluasi

Setelah model *Long Short-Term Memory* (LSTM) dibangun dan dilatih, tahap selanjutnya dalam penelitian ini adalah evaluasi model untuk mengukur kinerja prediksi yang dihasilkan. Evaluasi bertujuan untuk menilai sejauh mana model dapat memprediksi kualitas udara dalam ruangan dengan akurat berdasarkan data historis. Dalam penelitian ini, evaluasi dilakukan dengan membandingkan hasil prediksi model dengan data aktual dari dataset validasi yang telah dipisahkan sebelumnya.

Metrik evaluasi utama yang digunakan dalam penelitian ini adalah Mean Absolute Error (MAPE). MAPE mengukur rata-rata kesalahan absolut relatif antara nilai prediksi dan nilai aktual, yang dinyatakan dalam persentase terhadap nilai aktual. Metrik ini sangat intuitif dan bersifat skala-bebas, sehingga memudahkan perbandingan akurasi model pada data dengan skala yang berbeda. Semakin kecil nilai MAPE, semakin baik model dalam memprediksi kualitas udara, karena nilai MAPE yang rendah menunjukkan deviasi prediksi yang lebih kecil dari nilai sesungguhnya. MAPE dianggap sebagai indikator utama dalam penelitian ini untuk menilai akurasi prediksi. Selain MAPE, evaluasi model juga dilakukan menggunakan tiga metrik tambahan, yaitu *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan *Mean Absolute Error* (MAE), yang memberikan wawasan lebih lengkap mengenai kesalahan prediksi model:

- MSE mengukur rata-rata kesalahan kuadrat antara nilai prediksi dan nilai aktual, dengan memberikan penalti lebih besar pada kesalahan ekstrem. Hal ini membantu dalam mendeteksi anomali prediksi.

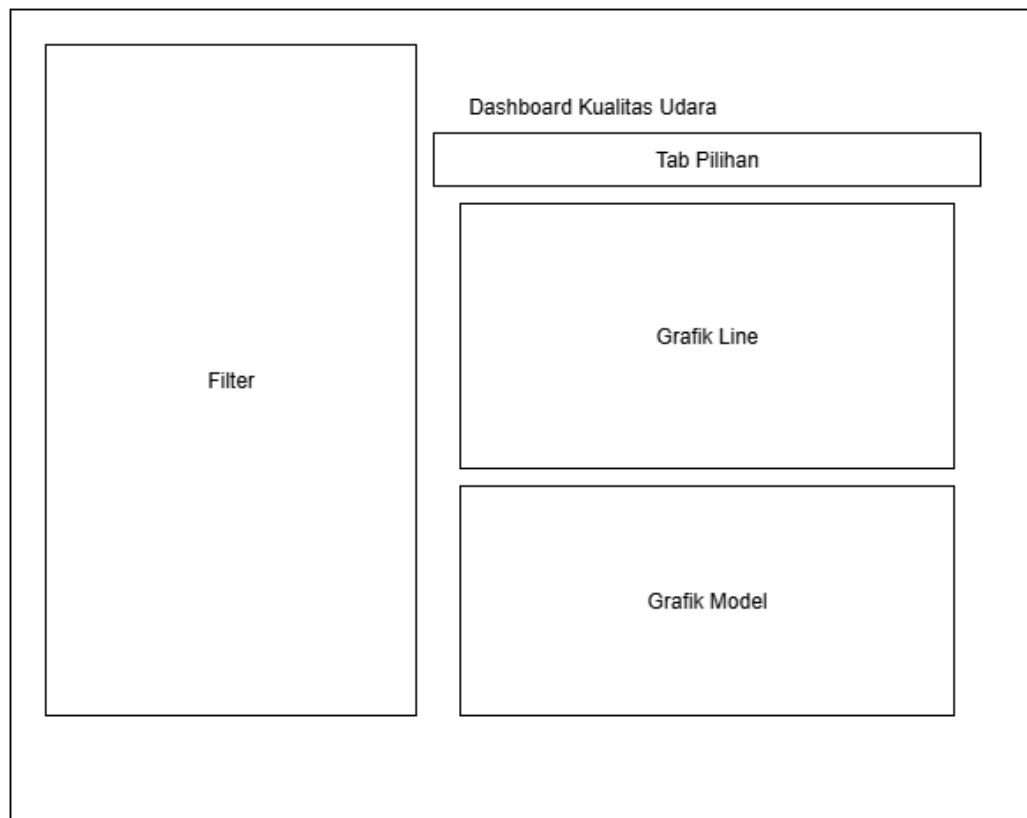
- RMSE merupakan akar kuadrat dari MSE, sehingga berada dalam satuan yang sama dengan data asli dan memudahkan interpretasi seberapa besar penyimpangan prediksi secara langsung.
- MAE menghitung rata-rata nilai absolut selisih antara prediksi dan aktual, memberikan gambaran umum besarnya kesalahan tanpa memperbesar kontribusi outlier.

Setelah pelatihan, model diuji pada data validasi, dan hasil prediksi dibandingkan dengan nilai aktual untuk menghitung keempat metrik tersebut. Model dengan nilai MSE, RMSE, MAE, dan MAPE yang rendah dianggap memiliki performa yang baik. Nilai MAPE yang rendah menjadi indikator utama dalam mengevaluasi kualitas prediksi model. Jika nilai MAPE masih tinggi, maka perlu dilakukan optimasi lebih lanjut, seperti penyesuaian arsitektur, atau perbaikan preprocessing data, agar rata-rata persentase kesalahan prediksi dapat ditekan.

Selain evaluasi kuantitatif, hasil prediksi juga divisualisasikan menggunakan grafik time series untuk membandingkan kurva antara nilai prediksi dan nilai aktual. Visualisasi ini memungkinkan analisis yang lebih intuitif terhadap pola prediksi model serta membantu dalam mengidentifikasi kemungkinan adanya tren atau pola yang belum ditangkap secara optimal oleh model. Untuk mendukung pengambilan keputusan berbasis data, hasil evaluasi juga ditampilkan dalam bentuk *dashboard* interaktif menggunakan Streamlit, yang mempermudah interpretasi dan analisis hasil prediksi.

3.3.6. Interpret

Setelah melakukan evaluasi model, tahap selanjutnya adalah memvisualisasikan hasil prediksi dan analisa dalam bentuk yang mudah dipahami dan interaktif. Untuk tujuan ini, penelitian ini menggunakan Streamlit untuk membuat *dashboard* interaktif yang menyajikan hasil prediksi kualitas udara dalam ruangan secara jelas dan memungkinkan pengguna untuk berinteraksi langsung dengan data.



Gambar 16. Mock Up *Dashboard* Kualitas Udara

Berdasarkan Gambar 16 yang merupakan mock up *dashboard* dimana pada bagian kiri *dashboard*, terdapat filter yang memungkinkan pengguna untuk mengeksplorasi hasil berdasarkan matriks matriks tertentu. Pengguna dapat memilih filter berikut:

- Nilai Sekolah: Pengguna dapat memilih data berdasarkan jenis sekolah yang dianalisis untuk membandingkan hasil prediksi kualitas udara antar sekolah yang berbeda.
- Season (Musim): Filter ini memungkinkan pengguna untuk melihat perbandingan prediksi kualitas udara berdasarkan musim, untuk menganalisis pengaruh faktor musiman terhadap kualitas udara dalam ruangan.
- Adanya Kipas di Langit-Langit: Filter ini memungkinkan pengguna untuk membandingkan data antara ruangan yang memiliki kipas langit-langit dan

yang tidak, untuk mengevaluasi dampaknya terhadap kualitas udara seperti sirkulasi udara dan konsentrasi CO₂.

Dashboard ini dilengkapi dengan tab pilihan, masing-masing dengan analisis berbeda terkait kualitas udara dalam ruangan sehingga user dapat memilih apa yang ingin di lihat berdasarkan kebutuhannya. Dibawah tab pilihan menunjukan grafik yang memberikan informasi serta pola yang dapat dilihat oleh user

Dengan menggunakan Streamlit, hasil prediksi tidak hanya divisualisasikan dengan cara yang lebih interaktif, tetapi juga memberikan wawasan lebih dalam mengenai efektivitas model dalam menangkap pola kualitas udara. Dengan demikian, deployment ini memungkinkan hasil penelitian diterapkan secara lebih luas dan aplikatif dalam bidang pemantauan serta prediksi kualitas udara dalam ruangan.

V. KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan penelitian yang telah saya lakukan, berikut adalah kesimpulan dari penelitian

1. Penelitian ini telah berhasil melakukan pemrosesan data kualitas udara dalam ruangan dan mengidentifikasi pola tren sehingga data tersebut telah memenuhi 4 kriteria kualitas data. Penelitian juga berhasil mengidentifikasi pola tren yang mana ruangan *permanent* lebih stabil dibandingkan *temporary*, dengan simpangan baku suhu masing-masing 2.039°C dan 4.056°C serta peningkatan kestabilan 49.73%. Konsentrasi CO₂ di ruang kelas mengikuti pola aktivitas harian pada setiap sekolah, sedangkan konsentrasi *Particulate Matter* (PM) tetap berada dalam ambang batas aman.
2. Model LSTM yang dilatih menunjukkan performa baik dalam memprediksi konsentrasi CO₂ di ruang kelas. Pada ruangan *permanent* (Sekolah-1), model memperoleh MAPE 6,9%, MSE 9.925,15 ppm², MAE 35,39 ppm, serta RMSE 99,6 ppm, sedangkan pada model pada ruangan *temporary* (Sekolah-2) diperoleh MAPE 11,16%, MSE 25.665,69 ppm², MAE 62,77 ppm, dan RMSE 160,2 ppm. Model dapat memprediksi dengan kesalahan relatif kecil.
3. *Dashboard* berhasil dikembangkan menggunakan Streamlit dan menampilkan hasil prediksi kualitas udara dalam ruangan dengan fitur utama, antara lain filter data, navigasi melalui tab serta kontrol legenda interaktif yang memudahkan pengguna untuk mengeksplorasi dan memahami data kualitas udara dalam ruangan.

5.2.Saran

Saran yang dapat saya sampaikan dalam penelitian dikemudian hari adalah

1. Pengumpulan data sebaiknya diperluas dengan memperhatikan keseimbangan distribusi data dari berbagai ruang dan perangkat. Data yang representatif dan berimbang akan meningkatkan kemampuan model dalam mengenali pola kualitas udara yang beragam, sehingga hasil prediksi menjadi lebih robust dan akurat.
2. Untuk meningkatkan akurasi prediksi kualitas udara, disarankan mengembangkan model hybrid yang menggabungkan LSTM dengan teknik pemodelan lain. Pendekatan hybrid ini berpotensi menangkap pola temporal dan spasial secara lebih efektif, serta mengurangi kesalahan prediksi terutama pada lonjakan episodik CO₂.
3. Apabila *dashboard* ingin diimplementasikan pada konteks universitas atau lingkungan lain, diperlukan penyesuaian struktur data dengan menambahkan parameter seperti fakultas, jurusan, gedung, dan jenis ruangan (misalnya ruang kuliah, laboratorium, atau perpustakaan) agar analisis kualitas udara tetap relevan dan akurat.

DAFTAR PUSTAKA

- [1] V. V. Tran, D. Park, and Y.-C. Lee, 'Indoor Air Pollution, Related Human Diseases, and Recent Trends in the Control and Improvement of Indoor Air Quality', *Int. J. Environ. Res. Public. Health*, vol. 17, no. 8, 2020, doi: 10.3390/ijerph17082927.
- [2] H. Y. W. Cheung *et al.*, 'Monitoring of indoor air quality at a large sailing cruise ship to assess ventilation performance and disease transmission risk', *Sci. Total Environ.*, vol. 962, p. 19, Jan. 2025, doi: 10.1016/j.scitotenv.2024.178286.
- [3] B. Talarosha, 'Jendela dan Dampaknya terhadap Konsentrasi CO₂ di dalam Ruang Kelas, Kajian Literatur', *J. Lingkung. Binaan Indones.*, vol. 6, no. 4, pp. 188–195, Dec. 2017, doi: 10.32315/jlbi.7.1.46.
- [4] J. Saini, M. Dutta, and G. Marques, 'Indoor Air Quality Monitoring Systems Based on Internet of Things: A Systematic Review', *Int. J. Environ. Res. Public. Health*, vol. 17, no. 14, p. 4942, Jul. 2020, doi: 10.3390/ijerph17144942.
- [5] H. Budianto and B. Sumanto, 'Perancangan Sistem Monitoring Kualitas Udara dalam Ruangan Berbasis Internet of Things', *J. List. Instrumentasi Dan Elektron. Terap.*, vol. 5, no. 1, p. 9, Apr. 2024, doi: 10.22146/juliet.v5i1.87423.
- [6] E. Z. Faesyari and Aprilana, 'Visualisasi Kualitas Udara Berdasarkan Parameter So₂ dan No₂ pada Kawasan Terbangun di Kabupaten Bekasi', *Pros. FTSP Ser.*, no. 2023: Prosiding FTSP Series 5, pp. 266–271, May 2023.
- [7] Y. Fan, X. Cao, J. Zhang, D. Lai, and L. Pang, 'Short-term exposure to indoor carbon dioxide and cognitive task performance: A systematic review and meta-analysis', *Build. Environ.*, vol. 237, p. 110331, Jun. 2023, doi: 10.1016/j.buildenv.2023.110331.

- [8] William S. and Cleveland, *Visualizing data*. Murray Hill, N.J.: At&T Bell Laboratories ; Published by Hobart Press, 1993.
- [9] F. J. Thendean, P. E. D. Tejokeosumo, and A. Rakhmawati, ‘Kajian Indoor Air Quality pada Rumah Tradisional Baileo Pegunungan di Seram Bagian Barat, Maluku’, *Intra*, vol. 7, no. 2, Art. no. 2, Jul. 2019.
- [10] S. Dimitroulopoulou *et al.*, ‘Indoor air quality guidelines from across the world: An appraisal considering energy saving, health, productivity, and comfort’, *Environ. Int.*, vol. 178, p. 108127, Aug. 2023, doi: 10.1016/j.envint.2023.108127.
- [11] G. Adamkiewicz and World Health Organization, Eds., *WHO guidelines for indoor air quality: selected pollutants*. Copenhagen, Denmark: World Health Organization, Regional Office for Europe, 2010.
- [12] J. G. Allen, P. MacNaughton, U. Satish, S. Santanam, J. Vallarino, and J. D. Spengler, ‘Associations of Cognitive Function Scores with Carbon Dioxide, Ventilation, and Volatile Organic Compound Exposures in Office Workers: A Controlled Exposure Study of Green and Conventional Office Environments’, *Environ. Health Perspect.*, vol. 124, no. 6, pp. 805–812, Jun. 2016, doi: 10.1289/ehp.1510037.
- [13] C. A. Pope and D. W. Dockery, ‘Health Effects of Fine Particulate Air Pollution: Lines that Connect’, *J. Air Waste Manag. Assoc.*, vol. 56, no. 6, pp. 709–742, Jun. 2006, doi: 10.1080/10473289.2006.10464485.
- [14] World Health Organization, *WHO global air quality guidelines: particulate matter (PM_{2.5} and PM₁₀), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*, 1 vols. Geneva: World Health Organization, 2021. Accessed: Sep. 15, 2025. [Online]. Available: <https://www.who.int/publications/i/item/9789240034228>
- [15] P. Wolkoff, K. Azuma, and P. Carrer, ‘Health, work performance, and risk of infection in office-like environments: The role of indoor temperature, air humidity, and ventilation’, *Int. J. Hyg. Environ. Health*, vol. 233, p. 113709, Apr. 2021, doi: 10.1016/j.ijheh.2021.113709.
- [16] Eli Trisnowati, Desika Rosiana Putri, Sabilla Safa Annisa Qurrota, Filda Khoirun Nikmah, and Danysa Mulyaningrum, ‘Analisis Konsep

- Termodinamika pada Produksi Kerupuk Sebagai Bentuk Kearifan Lokal di Magelang Jawa Tengah', *J. Pendidik. MIPA*, vol. 13, no. 1, pp. 268–273, Mar. 2023, doi: 10.37630/jpm.v13i1.795.
- [17] L. Cao, 'Data Science: A Comprehensive Overview', *ACM Comput. Surv.*, vol. 50, no. 3, pp. 1–42, May 2018, doi: 10.1145/3076253.
- [18] S. S. Abdul-Jabbar and A. K. Farhan, 'Data Analytics and Techniques: A Review', *ARO- Sci. J. KOYA Univ.*, vol. 10, no. 2, pp. 45–55, Oct. 2022, doi: 10.14500/aro.10975.
- [19] D. Tosi, R. Kokaj, and M. Roccetti, '15 years of Big Data: a systematic literature review', *J. Big Data*, vol. 11, no. 1, p. 73, May 2024, doi: 10.1186/s40537-024-00914-9.
- [20] K. Lepenioti, A. Bousdekis, D. Apostolou, and G. Mentzas, 'Prescriptive analytics: Literature review and research challenges', *Int. J. Inf. Manag.*, vol. 50, pp. 57–70, Feb. 2020, doi: 10.1016/j.ijinfomgt.2019.04.003.
- [21] R. Verma, 'Revolutionizing The Machine Learning Lifecycle With Cuttingedge Data Engineering', *Int. Res. J. Mod. Eng. Technol. Sci.*, vol. 4, no. 11, pp. 2341–2344, Nov. 2024.
- [22] R. Y. Wang, V. C. Storey, and C. P. Firth, 'A framework for analysis of data quality research', *IEEE Trans. Knowl. Data Eng.*, vol. 7, no. 4, pp. 623–640, Aug. 1995, doi: 10.1109/69.404034.
- [23] C. S. Descovi, A. C. Zuffo, S. M. Mohammadizadeh, L. F. Murillo-Bermúdez, and D. A. Sierra, 'Utilizing Long Short-Term Memory (Lstm) Networks For River Flow Prediction In The Brazilian Pantanal Basin', vol. 5, no. 39, pp. 1–15, Dec. 2023, doi: 10.15628/holos.2023.16315.
- [24] I. Cholissodin, *AI, Machine Learning & Deep Learning*. Malang: Fakultas Ilmu Komputer, Universitas Brawijaya, 2020.
- [25] F. Chollet, *Deep Learning with Python, Second Edition*. Shelter Island, New York: Manning Publications Co. LLC, 2021.
- [26] I Gusti Ayu Nandia Lestari and I Nyoman Dwi Arysna Mahendra, 'Prediksi Kualitas Udara dengan Menggunakan Metode Long Short-Term Memory dan Artificial Neural Network', *J. Sist. Dan Inform. JSI*, vol. 17, no. 2, pp. 121–129, May 2023, doi: 10.30864/jsi.v17i2.565.

- [27] M. Marwondo and T. Hidayah, 'Perbandingan Algoritma Long Short-Term Memory (LSTM) dan Gated Recurrent Unit (GRU) Untuk Prediksi Harga Emas Dunia', *Search*, vol. 21, no. 2, pp. 230–239, Apr. 2023, doi: 10.37278/insearch.v21i2.600.
- [28] S.-H. Noh, 'Analysis of Gradient Vanishing of RNNs and Performance Comparison', *Information*, vol. 12, no. 11, p. 442, Oct. 2021, doi: 10.3390/info12110442.
- [29] M. Gabriel and T. Auer, 'LSTM Deep Learning Models for Virtual Sensing of Indoor Air Pollutants: A Feasible Alternative to Physical Sensors', *Buildings*, vol. 13, no. 7, p. 1684, Jun. 2023, doi: 10.3390/buildings13071684.
- [30] R. Zaheer and H. Shaziya, 'A Study of the Optimization Algorithms in Deep Learning', in *2019 Third International Conference on Inventive Systems and Control (ICISC)*, Coimbatore, India: IEEE, Jan. 2019, pp. 536–539. doi: 10.1109/ICISC44355.2019.9036442.
- [31] K. Kumari, M. Bhardwaj, and S. Sharma, 'OSEMN Approach for Real Time Data Analysis', *Int. J. Eng. Manag. Res.*, vol. 10, no. 2, p. 4, 2020, doi: 10.18004/ijemr.v10i2.4. Available at SSRN: <https://ssrn.com/abstract=3584986> or <http://dx.doi.org/10.2139/ssrn.3584986>.
- [32] D. Leni, Muchlisinalahuddin, Maimuzar, Haris, and Hendra, 'Analisis Heatmap Korelasi dan Scatterplot untuk Mengidentifikasi Faktor-Faktor yang Mempengaruhi Pelabelan AC efisiensi Energi', *J. Rekayasa Mater. Manufaktur Dan Energi*, vol. 6, no. 1, Art. no. 1, Mar. 2023, doi: 10.30596/rmme.v6i1.13133.
- [33] S. Belouafa *et al.*, 'Statistical tools and approaches to validate analytical methods: methodology and practical examples★', *Int. J. Metrol. Qual. Eng.*, vol. 8, p. 9, 2017, doi: 10.1051/ijmqe/2016030.
- [34] M. I. J. Lamabelawa, 'Analisis Perhitungan Metode Interpolasi pada Data Time Series Kemiskinan di NTT', *J. Hoag – Teknol. Inf. STIKOM Uyelindo Kupang*, vol. 8, no. 1, pp. 640–646, 2018.
- [35] P. Kay S. De Guzman, A. Rafaella P. Honolario, N. I. Sore, and V. C. Quevedo, 'Evaluating the Accuracy of LSTM Forecasting Model and

- Intervention Programs in Combating Covid-19', in *Proceedings of the International Conference on Industrial Engineering and Operations Management*, Istanbul, Turkey: IEOM Society International, Mar. 2022, pp. 1887–1902. doi: 10.46254/an12.20220345.
- [36] D. R. Helsel and Hirsch, *Statistical methods in water resources*, vol. 49. in *Techniques and Methods*, vol. 49. Elsevier: U.S. Geological Survey, 1993.
- [37] I. Daqiqil, *Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python*, vol. 1. Unri Press, 2021.
- [38] N. Hewage and D. Meedeniya, 'Machine Learning Operations: A Survey on MLOps Tool Support', vol. 2, pp. 1–12, 2022, doi: <https://doi.org/10.48550/arXiv.2202.10169>.
- [39] R. G. Guntara, 'Deteksi Atap Bangunan Berbasis Citra Udara Menggunakan Google Colab dan Algoritma Deep Learning YOLOv7', *J. Manaj. Sist. Inf. JMASIF*, vol. 2, no. 1, Art. no. 1, May 2023, doi: 10.59431/jmasif.v2i1.156.
- [40] O. Ajayi, 'Exploratory Data Analytics of Air Pollutant Data for Air Quality Management Application', *Int. J. Technol. Syst.*, vol. 9, pp. 82–107, Jul. 2024, doi: 10.47604/ijts.2790.
- [41] A. Mota, C. Serôdio, A. Briga-Sá, and A. Valente, 'Implementation of an Internet of Things Architecture to Monitor Indoor Air Quality: A Case Study During Sleep Periods', *Sensors*, vol. 25, no. 6, p. 1683, Mar. 2025, doi: 10.3390/s25061683.
- [42] Y. Koçak and M. Koklu, 'Multi-layer long short-term memory (LSTM) prediction model on air pollution for Konya province', *Int. J. Appl. Math. Electron. Comput.*, vol. 10, no. 4, pp. 93–100, Dec. 2022, doi: 10.18100/ijamec.1208256.
- [43] L. Kristiana and D. Miyanto, 'Penambahan Parameter PM2.5 dalam Prediksi Kualitas Udara : Long Short Term Memory', *MIND Multimed. Artif. Intell. Netw. Database J.*, vol. 8, no. 2, Art. no. 2, Dec. 2023, doi: 10.26760/mindjournal.v8i2.188-202.
- [44] Y. Karyadi, 'Prediksi Kualitas Udara Dengan Metoda LSTM, Bidirectional LSTM, dan GRU', *JATISI J. Tek. Inform. Dan Sist. Inf.*, vol. 9, no. 1, pp. 671–684, Mar. 2022, doi: 10.35957/jatisi.v9i1.1588.

- [45] S. Fan, D. Hao, Y. Feng, K. Xia, and W. Yang, ‘A Hybrid Model for Air Quality Prediction Based on Data Decomposition’, *Information*, vol. 12, no. 5, p. 210, May 2021, doi: 10.3390/info12050210.
- [46] J. Cáceres-Tello and J. J. Galán-Hernández, ‘Analysis and Prediction of PM2.5 Pollution in Madrid: The Use of Prophet–Long Short-Term Memory Hybrid Models’, *AppliedMath*, vol. 4, no. 4, pp. 1428–1452, Nov. 2024, doi: 10.3390/appliedmath4040076.
- [47] Y. Zhu, S. A. Al-Ahmed, M. Z. Shakir, and J. I. Olszewska, ‘LSTM-Based IoT-Enabled CO2 Steady-State Forecasting for Indoor Air Quality Monitoring’, *Electronics*, vol. 12, no. 1, p. 107, Dec. 2022, doi: 10.3390/electronics12010107.
- [48] E. A. Gultom, K. D. S. Susilowati, and A. Kusmintarti, ‘Design of a Stock Forecasting Dashboard using Python-Streamlit and FB Prophet with AI’, *Formosa J. Sci. Technol.*, vol. 3, no. 11, pp. 2445–2464, Nov. 2024, doi: 10.55927/fjst.v3i11.12216.
- [49] Z. Zhang, H. Chen, and X. Huang, ‘Prediction of Air Quality Combining Wavelet Transform, DCCA Correlation Analysis and LSTM Model’, *Appl. Sci.*, vol. 13, no. 5, p. 2796, Feb. 2023, doi: 10.3390/app13052796.
- [50] R. E. Van Der Walt, S. S. Grobbelaar, and M. J. Booysen, ‘Dataset of indoor air parameter measurements relating to indoor air quality and thermal comfort in South African primary school classrooms of various building infrastructure types’, *Data Brief*, vol. 53, p. 110045, Apr. 2024, doi: 10.1016/j.dib.2024.110045.