

**IMPLEMENTASI ALGORITMA DEEP LEARNING UNTUK
MEMPREDIKSI KUALITAS UDARA DI PROVINSI LAMPUNG**

(Skripsi)

Oleh

BAGUS WAHYU PRATOMO

NPM 2015061003



**FAKULTAS TEKNIK
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2024**

**IMPLEMENTASI ALGORITMA DEEP LEARNING UNTUK
MEMPREDIKSI KUALITAS UDARA DI PROVINSI LAMPUNG**

Oleh

BAGUS WAHYU PRATOMO

NPM 2015061003

Skripsi

Sebagai Salah Satu Syarat Untuk Mencapai Gelar

SARJANA TEKNIK

Pada

Jurusan Teknik informatika

Fakultas Teknik Universitas Lampung



**FAKULTAS TEKNIK
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2024**

ABSTRAK

IMPLEMENTASI ALGORITMA DEEP LEARNING UNTUK MEMPREDIKSI KUALITAS UDARA DI PROVINSI LAMPUNG

Oleh

BAGUS WAHYU PRATOMO

Polusi udara ambien berdampak buruk pada kesehatan manusia. Indeks Kualitas Udara (AQI) digunakan untuk menilai kualitas udara. Analisis di 15 kota/kabupaten Provinsi Lampung menunjukkan AQI rata-rata tertinggi berada di daerah pinggiran kota, menjadikannya pilihan data utama untuk melatih model deep learning. Fitur PM2.5 dan PM10 ditemukan memiliki korelasi kuat dengan AQI, sehingga digunakan dalam prediksi AQI. Model RNN, LSTM, dan GRU dengan konfigurasi sederhana mencapai akurasi tinggi, dengan LSTM menunjukkan performa terbaik ($R^2 = 0,9360$; $MSE = 0,0011$). Model yang dilatih konsisten dengan indeks klasifikasi aktual dari Weatherbit.io, meski akurasi untuk daerah pinggiran kota lebih rendah dibanding daerah perkotaan. Studi ini menegaskan bahwa data historis dengan dua fitur polutan efektif memprediksi kualitas udara dan mendukung mitigasi polusi melalui pengoptimalan model.

Kata Kunci: Polusi Udara, Indeks Kualitas Udara, Deep Learning, RNN, LSTM, GRU, PM2.5, PM10

ABSTRACT

**IMPLEMENTATION OF DEEP LEARNING ALGORITHM
FOR PREDICTION AIR QUALITY INDEX
IN LAMPUNG PROVINCE**

By

BAGUS WAHYU PRATOMO

Ambient air pollution has adverse effects on human health. The Air Quality Index (AQI) is used to evaluate air quality. An analysis of 15 cities/regencies in Lampung Province reveals that suburban areas have the highest average AQI, making them the primary choice for training deep learning models. Features PM2.5 and PM10 show a strong correlation with AQI, making them ideal for AQI prediction. Sequential models like RNN, LSTM, and GRU with Single configurations achieved high accuracy, with LSTM performing best ($R^2 = 0.9360$; $MSE = 0.0011$). The models align with actual classification indices from Weatherbit.io, although suburban predictions were less accurate compared to urban areas. This study highlights that historical data with two pollutant features effectively predict air quality and support pollution mitigation efforts through model optimization.

Keywords: Air Pollution, Air Quality Index, Deep Learning, RNN, LSTM, GRU, PM2.5, PM10

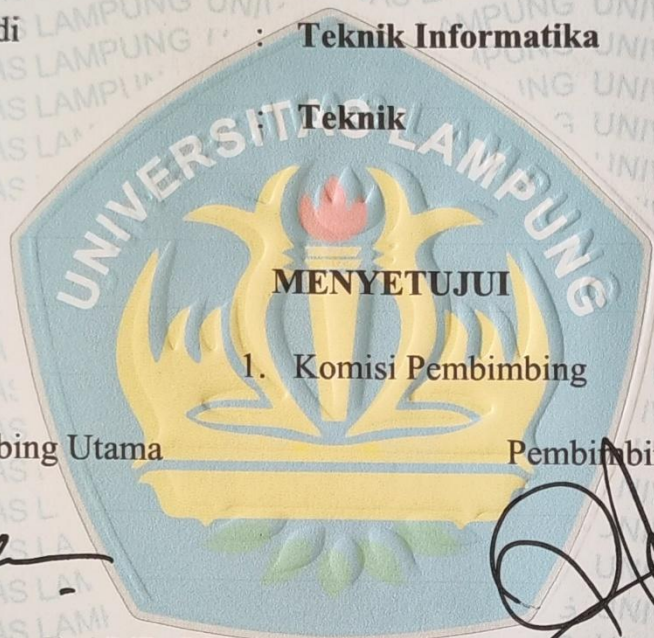
Judul Skripsi : **IMPLEMENTASI ALGORITMA DEEP LEARNING UNTUK MEMPREDIKSI KUALITAS UDARA DI PROVINSI LAMPUNG**

Nama Mahasiswa : **Bagus Wahyu Pratomo**

Nomor Pokok Mahasiswa : **2015061003**

Program Studi : **Teknik Informatika**

Fakultas : **Teknik**



1. **Komisi Pembimbing**

Pembimbing Utama

Pembimbing Pendamping

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

NIP. 196812071997031006

Puput Budi Wintoro, S.Kom., M.T.I.

NIP. 198410312019031004

2. **Mengetahui**

Ketua Jurusan

Ketua Program Studi

Herlinawati, S.T., M.T

NIP. 197103141999032001

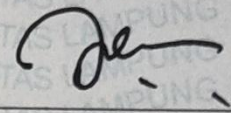
Yessi Mulyani, S.T., M.T.

NIP. 197312262000122001

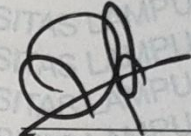
MENGESAHKAN

1. Tim Penguji

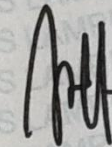
Ketua : Ir.M. Komarudin, S.T., M.T



Sekretaris : Puput Budi Wintoro, S.Kom., M.T.I.



Anggota : 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 : 13 Desember 2024

SURAT PERNYATAAN

Dengan ini, saya menyatakan bahwa skripsi dengan judul "Implementasi Algoritma Deep Learning untuk Memprediksi Kualitas Udara di Provinsi Lampung" adalah hasil karya saya sendiri. Segala hasil yang disajikan dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila di kemudian hari terbukti pernyataan ini tidak benar, saya bersedia menerima sanksi sesuai dengan ketentuan hukum yang berlaku.

Bandar Lampung, 18 Desember 2024
Pembuat Pernyataan,



Bagus Wahyu Pratomo
NPM. 2015061003

RIWAYAT HIDUP



Penulis lahir di Pringsewu pada 24 Juli 2002 sebagai anak pertama dari dua bersaudara, putra pasangan Supriyantomo dan Indah Sri Wahyuni. Pendidikan TK diselesaikan di 'Aisyiyah Bustanul Athfal, Pringsewu Timur, pada tahun 2008. Penulis melanjutkan pendidikan dasar di SDN 2 Pringsewu Timur dan lulus pada tahun 2014, kemudian menyelesaikan SMP di SMPN 2 Pringsewu pada tahun 2017 dan SMA di SMAN 1 Pringsewu pada tahun 2020. Pada tahun 2020, penulis diterima sebagai mahasiswa Teknik Informatika di Fakultas Teknik Universitas Lampung melalui jalur SNMPTN. Selama kuliah, penulis aktif sebagai asisten laboratorium komputer dan anggota bidang penelitian serta pengembangan di Himpunan Mahasiswa Teknik Elektro (HIMATRO). Pada tahun 2022, penulis meraih Beasiswa Bank Indonesia dan bergabung dengan komunitas Generasi Baru Indonesia (GenBI) Provinsi Lampung.

Di tahun 2023, penulis mengikuti program Kampus Merdeka Bank Indonesia Angkatan VI dan VII sebagai Data Analyst. Selain itu, penulis aktif dalam berbagai kegiatan informatika seperti Hackathon (Bio Farma X MIT Hacking Medicine, Healthkaton BPJS, GEMASTIK), pelatihan dan sertifikasi (HCIA Huawei), serta kompetisi karya tulis ilmiah (LKTI Fitopatologi).

MOTTO

"Setiap ilmu dapat dicapai melalui pengorbanan demi meraih keahlian dan kebolehan."

- Bagus Wahyu Pratomo

"Develop a passion for learning. If you do, you will never cease to grow."

- Anthony J. D'Angelo

"Apa yang menjadi milikmu akan kamu temukan dengan sendirinya"

- Ali bin Abi Thalib

PERSEMBAHAN

Alhamdulillahrabbi'l'alamin, segala puji bagi Allah SWT, Tuhan Yang Maha Esa, atas limpahan nikmat, rahmat, dan karunia-Nya sehingga skripsi ini dapat terselesaikan tepat pada waktunya. Karya ini kupersembahkan dengan penuh cinta kepada kedua orang tua tercinta, Supriyantomo dan Indah Sri Wahyuni, yang selalu memberikan doa, harapan, dan semangat tanpa henti. Terima kasih atas cinta, pengorbanan, dan dukungan yang tiada tara dalam setiap langkah perjalanan hidupku. Juga kepada seluruh keluarga besar, terima kasih atas motivasi dan dukungan yang senantiasa menjadi kekuatan bagiku.

Ucapan terima kasih juga kutujukan kepada diriku sendiri, atas perjuangan dan keteguhan hati dalam menempuh pendidikan ini. Terima kasih telah bertahan menghadapi segala tantangan dan tidak pernah menyerah untuk terus belajar. Semoga skripsi ini menjadi langkah awal yang membawa manfaat dan keberkahan bagi semua, serta menjadi wujud baktiku kepada keluarga dan lingkungan sekitarku.

SANWACANA

Alhamdulillah rabbil'alamin, segala puji dan syukur penulis panjatkan kepada Allah SWT atas rahmat, petunjuk, dan kasih sayang-Nya yang tiada henti, sehingga penulis dapat menyelesaikan skripsi ini dengan judul "**Implementasi Algoritma Deep Learning untuk Memprediksi Kualitas Udara di Provinsi Lampung**" tepat pada waktunya. Shalawat dan salam senantiasa tercurahkan kepada Baginda Nabi Muhammad SAW, semoga kita semua mendapat syafaat beliau di hari akhir nanti.

Skripsi ini tak akan terwujud tanpa dukungan, doa, dan peran berbagai pihak. Dengan penuh rasa hormat dan syukur, penulis mengucapkan terima kasih yang sebesar-besarnya kepada:

1. **Orang tua dan keluarga tercinta**, yang selalu memberikan doa, semangat, serta dukungan baik secara emosional maupun material. Kehadiran dan kasih sayang kalian adalah sumber kekuatan yang tiada habisnya.
2. **Bapak Dr. Eng. Helmy Fitriawan, S.T., M.Sc.**, selaku Dekan Fakultas Teknik Universitas Lampung, atas arahnya selama masa studi.
3. **Ibu Herlinawati, S.T., M.T.**, Ketua Jurusan Teknik Elektro Universitas Lampung, atas dukungannya terhadap kegiatan akademik.
4. **Ir. M. Komarudin, S.T., M.T.**, pembimbing utama yang dengan sabar membimbing dan memberikan arahan selama proses penelitian ini. Terima kasih atas waktu, kesabaran, dan nasihat berharga yang telah diberikan.
5. **Bapak Puput Budi Wintoro, S.Kom., M.T.I.**, pembimbing pendamping, yang tak henti-hentinya memberikan masukan dan membantu penulis menyempurnakan penelitian ini.

6. **Ibu Yessi Mulyani, S.T., M.T.**, Ketua Program Studi Teknik Informatika Universitas Lampung dan dosen penguji, yang telah memberikan kritik, saran, serta inspirasi selama proses penelitian maupun perkuliahan.
7. **Ibu Dr. Eng. Mardiana, S.T., M.T.**, pembimbing akademik yang telah dengan tulus memberikan masukan berharga selama perjalanan akademik penulis.
8. **Bapak Bayu Tejo Kartiko, S.P.**, paman yang selalu memberikan dukungan dan motivasi sehingga penulis dapat menyelesaikan pendidikan.
9. **Teman-teman, kakak, dan adik tingkat**, yang selalu hadir memberi semangat, dukungan, dan kebersamaan yang menguatkan selama perjalanan akademik ini.

Penulis menyadari bahwa skripsi ini jauh dari kata sempurna, baik dari segi isi maupun pengembangannya. Oleh karena itu, penulis dengan tulus menerima kritik dan saran yang membangun demi perbaikan di masa mendatang. Semoga hasil penelitian ini dapat memberikan manfaat, tidak hanya bagi pembaca, tetapi juga bagi perkembangan ilmu pengetahuan di bidang terkait.

Akhir kata, terima kasih yang mendalam juga penulis sampaikan kepada diri sendiri, atas perjuangan, ketekunan, dan kesabaran dalam menyelesaikan setiap tantangan selama perjalanan ini. Semoga langkah ini menjadi awal dari kontribusi yang lebih besar di masa depan.

Bandar Lampung, 18 Desember 2024

Penulis,

Bagus Wahyu Pratomo

DAFTAR ISI

	Halaman
DAFTAR TABEL	iii
DAFTAR GAMBAR.....	iv
I. PENDAHULUAN.....	1
I.1 Latar Belakang.....	1
I.2 Rumusan Masalah.....	6
I.3 Tujuan Penelitian	6
I.4 Manfaat Penelitian	6
I.5 Batasan Masalah	7
I.6 Sistematika Penulisan Skripsi.....	8
II. TINJAUAN PUSTAKA	9
2.1 Air Quality Index (AQI).....	9
2.1.1 Polutan Penyusun AQI.....	9
2.1.2 Metode Perhitungan Kualitas Udara	10
2.2 Provinsi Lampung	14
2.3 Peramalan (<i>Forecasting</i>).....	14
2.4 <i>Sequential Data</i> (Data Berurutan).....	14
2.5 Korelasi Pearson.....	15
2.6 Artificial Inteligence (AI)	16
2.7 Machine Learning	17
2.8 Deep learning	17
2.8.1 Jaringan Syaraf Tiruan (JST)	18
2.8.2 <i>Single Recurrent Neural Network</i> (RNN).....	20
2.8.3 Long short-term memory (LSTM)	24
2.8.4 Gated <i>Recurrent Unit</i> (GRU).....	28

2.9 Fitur <i>Dropout</i> dan Teknik <i>Bidirectional</i>	29
2.9.1 <i>Dropout</i>	29
2.9.2 Teknik <i>Bidirectional</i>	30
2.10 Fungsi Aktivasi	31
2.11 Index Performa.....	32
2.12 Cross-Industry Standard Process for Data Mining (<i>CRISP-DM</i>)	34
2.13 Flask	35
2.14 Penelitian Terkait	36
III. METODOLOGI PENELITIAN	46
3.1 Waktu dan Tempat	46
3.2 Alat dan Bahan.....	46
3.2.1 Alat.....	46
3.2.2 Bahan	48
3.3 Metode Penelitian.....	49
IV. HASIL DAN PEMBAHASAN.....	54
4.1 Pembahasan Penelitian.....	54
4.1.1 Pemahaman Bisnis	54
4.1.2 Pemahaman data terkait tren kualitas udara di Provinsi Lampung	55
4.1.3 Pemahaman Data terkait Pemodelan data latih.....	57
4.1.4 Persiapan Data.....	61
4.1.5 Pembuatan Model	65
4.1.6 Evaluasi	68
4.1.7 Deployment.....	73
4.2 Hasil Penelitian	74
4.1.1 Tren Kualitas udara di provinsi lampung.....	74
4.1.2 Model RNN, LSTM dan GRU yang optimal.....	82
4.1.3 Penerapan model pada data langsung menggunakan framework flask	88
V. KESIMPULAN DAN SARAN	114
5.1 Kesimpulan	114
5.2 Saran.....	115
DAFTAR PUSTAKA	116

DAFTAR TABEL

Tabel		Halaman
1	Tabel Konversi Nilai Konsentrasi Parameter AQI	11
2	Penelitian Terkait	36
3	Jadwal Kegiatan Penelitian	46
4	Software Pengembangan Deep learning.	47
5	Titik Lokasi Pengambilan Data	55
6	Explorasi Data Analisis AQI Lampung	56
7	Tabel Akurasi Model RNN	82
8	Tabel Akurasi Model LSTM.....	84
9	Tabel Akurasi Model GRU	86
10	Nilai MSE pada model live.....	112
11	Tingkat Akurasi R^2 model data live	113

DAFTAR GAMBAR

Gambar		Halaman
1	Kategori Kualitas Udara menurut EPA.....	13
2	ANN Satu Lapisan Umpan Maju (Single Layer feedforward network)	19
3	ANN Multilayer Umpan Maju (Multilayer feedforward network).....	19
4	Arsitektur ANN berulang (Recurrent Network) model Elman.....	20
5	Arsitektur ANN berulang (Recurrent Network) model Jordan.....	20
6	Arsitektur RNN.....	21
7	Sebuah recurrent neural network dapat dibuat lebih dalam dengan berbagai cara (Pascanu et al., 2014a).....	23
8	LSTM.....	25
9	Arsitektur CRISP-DM	34
10	Flowchart Penelitian	49
11	Scatter Plot AQI All Cities	56
12	Raw data pelatihan.....	57
13	Data sebelum dibersihkan	58
14	Data setelah dibersihkan	58
15	Dataset setelah dibersihkan.....	58
16	Line Plot Masing Masing Fitur dengan AQI dalam 1 minggu	59
17	Hubungan Seluruh Fitur dengan AQI dalam 1 minggu.....	60
18	Korelasi Pearson	60
19	Data Fitur	62
20	Data Target.....	62
21	Reshape target menjadi array 2D.....	62
22	Data Fitur Setelah di Normalisasi	63
23	Data Target Setelah di Normalisasi	63
24	Fungsi untuk membuat data menjadi berurutan.....	63
25	Data fitur menjadi array 3D	64

26	Data Target array 1D	64
27	Ukuran Data yang telah dibagi	65
28	Arsitektur RNN Double 50 neurons, 100 epochs, 0.001 lr.....	65
29	Arsitektur LSTM Double 100 neurons, 100 epochs, 0.0005 lr.....	66
30	Arsitektur GRU Bidirectional 100 neurons, 100 epochs, 0.001 lr.....	67
31	Grafik loss pelatihan model RNN.....	68
32	Tabel prediksi dan aktual nilai AQI model RNN	69
33	Grafik prediksi dan aktual nilai AQI model RNN.....	69
34	Grafik loss pelatihan model LSTM.....	70
35	Tabel prediksi dan aktual nilai AQI model LSTM	70
36	Grafik prediksi dan aktual nilai AQI model LSTM.....	71
37	Grafik loss pelatihan model GRU.....	71
38	Tabel prediksi dan aktual nilai AQI model GRU	72
39	Grafik prediksi dan aktual nilai AQI model GRU	72
40	Tren AQI Bandar Lampung	74
41	Tren AQI Lampung Barat.....	74
42	Tren AQI Lampung Utara.....	75
43	Tren AQI Lampung Selatan.....	75
44	Tren AQI Lampung Tengah.....	76
45	Tren AQI Lampung Timur.....	76
46	Tren AQI Mesuji.....	76
47	Tren AQI Metro	77
48	Tren AQI Pesawaran.....	77
49	Tren AQI Pesisir Barat.....	78
50	Tren AQI Pringsewu	78
51	Tren AQI Tanggamus	79
52	Tren AQI Tulang Bawang	79
53	Tren AQI Tulang Bawang Barat.....	79
54	Tren AQI Way Kanan.....	80
55	Tren di 15 kota/kabupaten di Provinsi Lampung.....	81
56	Dashboard Sederhana AQI Prediction	88
57	Prediksi AQI pada wilayah Bandar Lampung RNN.....	89

58	Prediksi AQI pada wilayah Lampung Barat RNN.....	89
59	Prediksi AQI pada wilayah Lampung Utara RNN	90
60	Prediksi AQI pada wilayah Lampung Selatan RNN.....	90
61	Prediksi AQI pada wilayah Lampung Tengah RNN	91
62	Prediksi AQI pada wilayah Lampung Timur RNN	91
63	Prediksi AQI pada wilayah Mesuji RNN.....	92
64	Prediksi AQI pada wilayah Metro RNN	92
65	Prediksi AQI pada wilayah Pesawaran RNN.....	93
66	Prediksi AQI pada wilayah Pesisir Barat RNN	93
67	Prediksi AQI pada wilayah Pringsewu RNN.....	94
68	Prediksi AQI pada wilayah Tanggamus RNN	95
69	Prediksi AQI pada wilayah Tulang Bawang RNN	95
70	Prediksi AQI pada wilayah Tulang Bawang Barat	96
71	Prediksi AQI pada wilayah Way Kanan RNN.....	96
72	Prediksi AQI pada wilayah Bandar Lampung LSTM.....	97
73	Prediksi AQI pada wilayah Lampung Barat LSTM	97
74	Prediksi AQI pada wilayah Lampung Utara LSTM	98
75	Prediksi AQI pada wilayah Lampung Selatan LSTM	98
76	Prediksi AQI pada wilayah Lampung Tengah LSTM.....	99
77	Prediksi AQI pada wilayah Lampung Timur LSTM.....	99
78	Prediksi AQI pada wilayah Mesuji LSTM	100
79	Prediksi AQI pada wilayah Metro LSTM.....	100
80	Prediksi AQI pada wilayah Pesawaran LSTM	101
81	Prediksi AQI pada wilayah Pesisir Barat LSTM.....	101
82	Prediksi AQI pada wilayah Pringsewu LSTM.....	102
83	Prediksi AQI pada wilayah Tanggamus LSTM.....	103
84	Prediksi AQI pada wilayah Tulang Bawang LSTM	103
85	Prediksi AQI pada wilayah Tulang Bawang Barat LSTM	104
86	Prediksi AQI pada wilayah Way Kanan LSTM	104
87	Prediksi AQI pada wilayah Bandar Lampung GRU.....	105
88	Prediksi AQI pada wilayah Lampung Barat GRU.....	105
89	Prediksi AQI pada wilayah Lampung Utara GRU	106

90	Prediksi AQI pada wilayah Lampung Selatan GRU.....	106
91	Prediksi AQI pada wilayah Lampung tengah GRU.....	107
92	Prediksi AQI pada wilayah Lampung Timur GRU	107
93	Prediksi AQI pada wilayah Mesuji GRU.....	108
94	Prediksi AQI pada wilayah Metro GRU	108
95	Prediksi AQI pada wilayah Pesawaran GRU.....	109
96	Prediksi AQI pada wilayah Pesisir Barat GRU	109
97	Prediksi AQI pada wilayah Pringsewu GRU.....	110
98	Prediksi AQI pada wilayah pada wilayah Tanggamus GRU.....	110
99	Prediksi AQI pada wilayah Tulang Bawang GRU	111
100	Prediksi AQI pada wilayah Tulang bawang barat GRU	111
101	Prediksi AQI pada wilayah Way Kanan GRU.....	112

I. PENDAHULUAN

I.1 Latar Belakang

Menurut IQAir, 7 juta orang meninggal dikarenakan polusi udara setiap tahun dan miliaran orang menderita akibat kualitas udara yang buruk. Masyarakat banyak tidak memiliki akses ke informasi kualitas udara yang tepat waktu, bahkan polusi udara tidak diketahui. Pada bulan Agustus tahun 2023, Indonesia dihadapkan pada tantangan serius terkait kualitas udara, khususnya di Ibukota, Jakarta, yang mencapai peringkat tertinggi sebagai kota dengan kualitas udara terburuk di dunia. Informasi ini terungkap pada tanggal 10 Agustus 2023, ketika IQAir mengeluarkan data yang menunjukkan bahwa konsentrasi $PM_{2.5}$, sebagai salah satu polutan utama udara, mencapai 158 mikrogram per meter kubik standar. Angka ini mencerminkan peningkatan yang signifikan sebanyak 11.6 kali lipat dibandingkan dengan nilai panduan tahunan Kualitas Udara yang telah ditetapkan oleh Organisasi Kesehatan Dunia (WHO) (CNBC Indonesia, 2023).

Dalam suatu kajian yang dilakukan oleh Puji Lestari, seorang Peneliti dan Guru Besar Teknik Lingkungan di Institut Teknologi Bandung (ITB), menjelaskan bahwa dari tahun 2017 hingga 2023, puncak polusi udara di Jakarta terjadi pada bulan Juli-Agustus. Selanjutnya, konsentrasi polusi menurun selama musim penghujan, hanya untuk kembali meningkat saat musim kemarau. Kualitas udara buruk membawa dampak kesehatan serius, $PM_{2.5}$ terserap ke dalam tubuh dapat menyebabkan bronkitis, pneumonia, penyakit paru obstruktif kronik, jantung penyakit, stroke, dan penyakit pernapasan[1].

Menurut Dede Anwar Musadad, Peneliti Ahli Utama di Pusat Riset Kesehatan Masyarakat dan Gizi Badan Riset dan Inovasi Nasional (BRIN), lima penyakit utama akibat polusi udara di Indonesia pada tahun 2019 adalah stroke, penyakit jantung iskemik, diabetes melitus, penyakit paru obstruktif kronis (PPOK), dan

gangguan neonatal. Data dari Organisasi Kesehatan Dunia (WHO) menunjukkan bahwa pada tahun 2020, polusi udara dalam ruangan diperkirakan menyebabkan 3,2 juta kematian global per tahun. Angka ini sangat mengkhawatirkan, terutama karena lebih dari 237.000 kematian terjadi pada anak-anak balita. Selain itu, polusi udara dalam ruangan merupakan faktor risiko utama untuk penyakit tidak menular serius seperti stroke, penyakit paru obstruktif kronis, penyakit jantung iskemik, dan kanker paru-paru (BRIN, 2024). Sedangkan menurut Riskesdas, pada tahun 2018, total pasien penyakit pernafasan (ISPA, Tb Paru dan Asma) mencapai 22.160 pasien di pedesaan dan 9.302 pasien di perkotaan pada tahun 2018 di provinsi Lampung[2].

Hal ini menyoroti pentingnya pemantauan rutin terhadap polusi udara, baik polusi udara luar ruangan (*ambient air pollution*) maupun dalam ruangan (*household air pollution*). Pengawasan kualitas udara dalam ruangan harus dilakukan oleh petugas puskesmas, sebagaimana diatur dalam Permenkes No. 1077 Tahun 2011 tentang Pedoman Penyehatan Udara dalam Ruang Rumah.

Indeks Kualitas Udara (IKU) Provinsi Lampung disusun melalui pemantauan kualitas udara menggunakan metode *Passive Sampler* di empat lokasi *strategis*, termasuk area transportasi, industri, pemukiman, dan perkantoran/perdagangan. Pemantauan dilakukan dua kali dalam setahun untuk mendapatkan data yang representatif. Metodologi perhitungan IKU mengadopsi Program European Union melalui *European Regional Development Fund*, dengan menggunakan *Common Air Quality Index (CAQI)* dari proyek "*Common Information to European Air*" (*Citeair II*) tahun 2012. CAQI telah digunakan sejak tahun 2006 dan dapat diakses melalui situs web www.airqualitynow.eu. Indeks ini dihitung berdasarkan data rata-rata per jam, harian, dan tahunan. Cara perhitungan indeksnya adalah dengan membandingkan nilai rata-rata tahunan dengan standar yang ditetapkan dalam *EU Directives*. Jika nilai indeks IKU lebih dari 1, maka kualitas udara melebihi standar Uni Eropa. Sebaliknya, jika nilai indeks IKU sama dengan atau kurang dari 1, maka kualitas udara memenuhi atau lebih baik dari standar Uni Eropa[3].

Tidak hanya CAQI, di dunia memiliki beberapa metode perhitungan kualitas udara. Di Kanada, meski sudah lama menggunakan indeks kualitas udara provinsi

provincial air quality indices (AQIs), fokusnya lebih pada pencapaian emisi rendah dibanding kesehatan manusia. Untuk itu, Kanada menggunakan *Air Quality Health Index (AQHI)*. *AQHI* memiliki skala 1-10+ dimana semakin tinggi angkanya semakin besar risiko kesehatan. Hong Kong juga menggunakan *AQHI* sejak 2013 untuk menggantikan indeks polusi udara sebelumnya. Eropa tidak hanya memiliki *Common Air Quality Index (CAQI)* yang sudah digunakan sejak 2006, tetapi juga memiliki indeks yang lebih baru. Pada November 2017, Badan Lingkungan Eropa mengumumkan *European Air Quality Index (EAQI)* dan mendorong penggunaannya untuk menginformasikan kualitas udara kepada masyarakat[4]. Singapore menggunakan *Pollutant Standards Index (PSI)* untuk melaporkan kualitas udaranya. Kemudian *Environmental Protection Agency (EPA)* di Amerika Serikat membuat sebuah *Air Quality Index (AQI)* untuk melaporkan kualitas udara disana. *AQI* dibagi menjadi enam kategori yang menunjukkan peningkatan risiko kesehatan. Nilai *AQI* di atas 300 berarti kualitas udara berbahaya, sementara di bawah 50 menunjukkan kualitas udara yang baik.

Di Indonesia, Indeks Kualitas Udara (IKU) disebut juga dengan Indeks Standar Pencemar Udara (ISPU). ISPU, atau Indeks Standar Pencemar Udara, adalah angka tanpa satuan yang menunjukkan tingkat kualitas udara di suatu tempat. Di daerah rawan kebakaran hutan dan lahan, ISPU dapat digunakan sebagai sistem peringatan dini bagi masyarakat sekitar. Kementerian Lingkungan Hidup dan Kehutanan (KLHK) mengolah indeks kualitas lingkungan hidup sebagai bahan informasi untuk mendukung proses pengambilan kebijakan yang berkaitan dengan perlindungan dan pengelolaan lingkungan hidup. Salah satu komponen di dalam IKLH adalah kualitas udara.

Pengelolaan IKU memerlukan pemantauan berkelanjutan dan perkiraan konsentrasi polutan yang akurat. Stasiun Pemantau Kualitas Udara Ambien (SPKUA) telah banyak melakukan pemantauan yang menghasilkan banyak data historis. Metode pengukuran kualitas udara ambien meliputi metode otomatis dan metode manual. Metode kontinyu otomatis menggunakan *air quality monitoring system (AQMS)*. Metode otomatis ini meliputi permanen (*fixed station*) dan bergerak (*mobile station*). Sedangkan metode manual menggunakan metode manual aktif dan metode manual pasif (*passive sampler*). Sesuai lampiran III PermenLH no. 12 tahun 2010,

Metode manual dilakukan dengan cara pengambilan sampel udara terlebih dahulu lalu dianalisis di laboratorium. Metode manual ini dibedakan lagi menjadi metode passive dan aktif. Perbedaan ini didasarkan pada ada tidaknya pompa untuk mengambil sampel udara. Metode otomatis dilakukan dengan menggunakan alat yang dapat mengukur kualitas udara secara langsung sekaligus menyimpan datanya[5].

Metode numerik sangat berguna untuk memodelkan kualitas udara dengan resolusi spasial dan temporal yang terperinci serta pemodelan kimia dan fisik yang kompleks[6]. *Deep learning* adalah metode yang menggunakan jaringan syaraf tiruan yang kompleks. Jaringan syaraf tiruan adalah salah satu metode kecerdasan buatan yang memanfaatkan data sebagai inti dalam pembelajarannya. Salah satu penerapannya adalah peramalan (*forecasting*). Peramalan kualitas udara dapat menjadi landasan bagi pembuat kebijakan dalam menyusun rencana jangka pendek, menengah, dan jangka panjang untuk mengelola dan mengurangi dampak pencemaran udara. Hal ini memungkinkan pemerintah dan pemangku kepentingan untuk mengambil langkah-langkah pencegahan dan perlindungan yang sesuai sebelum terjadinya peningkatan polusi udara yang dapat berdampak buruk bagi masyarakat. Secara sederhana, metode yang umum digunakan untuk meramalkan kualitas udara terbagi dalam tiga kategori utama: pendekatan **klimatologi**, **metode statistik** seperti Pohon Klasifikasi dan Regresi (CART), analisis regresi, dan jaringan saraf (neural network), serta **model tiga dimensi** yang mencakup model emisi, meteorologi, dan kimia. Penelitian ini menggunakan algoritma deep learning, yang termasuk dalam penerapan metode statistik (AS EPA, 1999).

Penelitian terkait prediksi kualitas udara memiliki banyak metode. Jingyang Wang, dkk. dalam penelitiannya menggunakan metode [3]*CNN dan AGU (Attention Gate Units)* yang ditujukan untuk mengatasi masalah *vanishing gradient* dan *exploding gradient* dari *Recurrent Neural Network (RNN)*[7]. Dalam penelitian yang dilakukan oleh Benny Khalid Hidayatullah, dkk. Meneliti terkait kualitas udara di kota jakarta pada tahun 2022. Penelitian tersebut menggunakan LSTM untuk memprediksi polutan ISPU (PM_{10} , SO_2 , NO_2 , CO , O_3) dengan hasil pengujian parameter terbaik dari kelima zat tersebut menghasilkan rentang nilai $RMSE$ dari 0.0431 sampai 0.07237 dan skor R^2 dari rentang nilai 0.58157 hingga 0.88621[8].

Selain itu, Ali khumaidi, dkk. meneliti pemodelan LSTM untuk memprediksi kualitas udara di kota Bandung pada tahun 2022, dengan menggunakan parameter PM_{10} , ISPU, suhu, dan kelembaban. Prediksi ISPU juga menunjukkan kinerja baik dengan *RMSE Train Score* 1,89 dan *Test Score* 1,85[9]. Dalam konteks lainnya, LSTM dapat untuk klasifikasi Pada Hate Speech[10], Memprediksi Harga Bahan Pokok Nasional[11] dan *Prediction of Water Levels* [12].

Dari berbagai penelitian tersebut, Model Long Short-Term Memory (LSTM), *Recurrent Neural Network* (RNN), dan *Gated Recurrent Unit* (GRU) sangat cocok untuk prediksi karena kemampuannya dalam menangkap dan memahami pola-pola kompleks dalam data sekuensial. Berbeda dengan model tradisional, LSTM dan GRU memiliki kemampuan untuk mempertahankan informasi jangka panjang, sehingga sangat efektif dalam memodelkan data dengan konteks temporal yang penting. Fitur utama dari LSTM adalah gate yang memungkinkan mereka untuk memilih informasi yang relevan untuk dipelajari, dipertahankan, atau dilupakan, sedangkan GRU menyederhanakan mekanisme ini dengan menggunakan dua jenis gate. RNN pada umumnya, termasuk LSTM dan GRU, memungkinkan model untuk menangkap hubungan jangka panjang dalam data. Oleh karena itu, RNN, LSTM, dan GRU sering digunakan dalam berbagai aplikasi prediksi, termasuk dalam pemodelan bahasa alami, peramalan deret waktu, dan analisis sekuensial lainnya, karena kemampuannya yang unggul dalam memprediksi berbagai fenomena yang melibatkan urutan data.[13].

KDD (Knowledge Discovery in Databases), *OSEMN (Obtain, Scrub, Explore, Model, and iNterpret)*, dan *CRISP-DM (Cross-Industry Standard Process for Data Mining)* adalah tiga metodologi populer dalam data mining dan analisis data. *KDD* mencakup proses menyeluruh dari ekstraksi pengetahuan dari data, yang meliputi tahap seleksi, praproses, transformasi, penambangan data, dan interpretasi/evaluasi. *OSEMN* adalah pendekatan siklus yang terdiri dari lima langkah: memperoleh data, membersihkan data, mengeksplorasi data, memodelkan data, dan menginterpretasikan hasil. Sementara itu, *CRISP-DM* adalah model proses yang terdiri dari enam fase: pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan deployment. Masing-masing metodologi ini menyediakan kerangka kerja sistematis untuk melakukan analisis data yang efektif dan efisien.

Dalam penelitian ini, proses implementasi merupakan proses deployment, dimana dalam metodologi *CRISP-DM* cocok digunakan dalam penelitian ini. Selain itu, pada metodologi *KDD* memerlukan akses penuh ke server agar penelitian optimal dan mendapatkan evaluasi dari penambangan data. Sedangkan untuk metodologi *OSEMN* berakhir dengan evaluasi dari pemodelan sehingga tidak cocok karena belum mendapatkan kesimpulan dari implementasi model.

I.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini antara lain sebagai berikut.

1. Kota/kabupaten mana di Provinsi Lampung yang paling representatif sebagai sumber dataset untuk melatih model deep learning melalui eksplorasi data analisis?
2. Bagaimana tingkat akurasi dari model deep learning yang digunakan dalam pelatihan?
3. Bagaimana tingkat akurasi dari model deep learning yang diterapkan menggunakan data langsung dari ke-15 kota di Provinsi Lampung?

I.3 Tujuan Penelitian

Pada penelitian ini, terdapat beberapa tujuan penelitian sebagai berikut.

1. Menggunakan salah satu kota untuk dijadikan dataset dalam pembuatan model deep learning.
2. Mendapatkan akurasi dan kelebihan serta kekurangan dari model deep learning yang digunakan.
3. Mendapatkan akurasi dari model deep learning yang digunakan.

I.4 Manfaat Penelitian

1. Teoritis

Secara teoritis, penelitian ini bertujuan untuk berkontribusi dalam mendukung penelitian-penelitian lain, terutama yang berkaitan dengan kualitas udara dan penggunaan teknologi kecerdasan buatan seperti *deep learning* dengan berfokus pada penggunaan algoritma RNN, LSTM dan GRU dasar.

2. Praktis

a. Bagi Akademik

Bagi kalangan akademik, penelitian ini dapat menjadi sumber referensi untuk meningkatkan pemahaman dan pengembangan teori terkait penerapan model RNN, LSTM dan GRU dalam memonitor kualitas udara di provinsi lampung.

b. Bagi Masyarakat

Bagi masyarakat, hasil penelitian ini dapat memberikan informasi dan referensi tentang kondisi kualitas udara di Provinsi Lampung.

3. Kebijakan

Dari segi kebijakan, diharapkan hasil penelitian ini dapat menjadi inovasi bagi pemerintah dan lembaga terkait dalam merencanakan kebijakan terkait prediksi dan pemantauan kualitas udara, khususnya di wilayah Lampung.

I.5 Batasan Masalah

Dalam skripsi ini, berikut adalah batasan masalah yang dijelaskan dalam bentuk poin:

1. Target prediksi adalah pemantauan kualitas udara (*AQI*).
2. Data yang digunakan untuk pelatihan model merupakan data history pada salah satu kota di Provinsi Lampung. Data diperoleh menggunakan *API Weatherbit*. Menimbang keterbatasan dalam pengambilan data, maka tren yang dapat ditampilkan hanya 2 tahun terakhir.
3. Arsitektur yang digunakan merupakan arsitektur *Single*, *Double*, *Double* dengan penambahan regulasi *Dropout* dan *Bidirectional*. Hyperparameter yang digunakan adalah menggunakan timesteps 24 dikarenakan kebutuhan data PM_{25} dan PM_{10} selama 24 jam untuk menghitung *AQI*, *epochs* 50 dan

100, *learning rate* 0.001 dan 0.005, *neurons* 50 dan 100 dan regulasi *Dropout* 0.2.

4. Evaluasi kinerja model dilakukan dengan memeriksa indikator berikut:, *Mean Squared Error (MSE)* dan *Koefisien Determinasi (R^2)*.
5. Pengembangan model dan visualisasi hasil dilakukan menggunakan framework flask secara lokal menggunakan localhost untuk menampilkan prediksi wilayah lainnya.

I.6 Sistematika Penulisan Skripsi

Sistematika penulisan skripsi ini terdiri dari 5 (lima) bab sebagai berikut:

BAB 1: PENDAHULUAN

Memuat latar belakang penyusunan skripsi ini, perumusan masalah, tujuan penelitian, manfaat penelitian dan batasan masalah serta sistematika penulisan dalam penelitian ini.

BAB II : TINJAUAN PUSTAKA

Memuat penjelasan terkait penelitian terkait dan dasar teori yang terkait dengan penelitian ini.

BAB III : METODOLOGI PENELITIAN

Memuat waktu dan tempat penelitian, tahapan penelitian secara umum, dan tahapan penelitian secara terperinci. Metodologi yang digunakan dalam penelitian ini adalah *CRISP-DM*.

BAB IV : HASIL DAN PEMBAHASAN:

Memuat hasil penelitian dengan berdasarkan pada perumusan masalah serta pembahasan terkait hasil penelitian.

BAB V KESIMPULAN

Memuat kesimpulan dari analisa yang dilakukan pada penelitian ini.

DAFTAR PUSTAKA

II. TINJAUAN PUSTAKA

2.1 Air Quality Index (AQI)

Indeks Kualitas Udara (AQI) adalah metrik yang digunakan oleh *Environmental Protection Agency (EPA)* di Amerika Serikat dan lembaga serupa di negara lain untuk mengukur seberapa sehat udara yang kita hirup. Ini memberikan informasi tentang tingkat pencemaran udara dan efek kesehatannya bagi masyarakat [14].

2.1.1 Polutan Penyusun AQI

AQI mengukur beberapa polutan udara yang umum, seperti partikulat (PM_{10} dan $PM_{2.5}$), ozon, nitrogen dioksida, sulfur dioksida, dan karbon monoksida. EPA menetapkan nilai batas untuk masing-masing polutan ini, dan setiap nilai dipetakan ke dalam kategori kualitas udara: baik, sedang, tidak sehat bagi kelompok sensitif, tidak sehat, sangat tidak sehat, dan berbahaya. Berikut penjelasan Air Quality Index (EPA) meliputi parameter sebagai berikut.

a. $PM_{2.5}$

Particulate Matter 2.5 ($PM_{2.5}$) adalah partikel kecil dalam udara dengan diameter kurang dari 2.5 mikrometer[14]. Partikel ini dapat berasal dari berbagai sumber, termasuk kendaraan bermotor, industri, dan pembakaran biomassa. $PM_{2.5}$ dapat masuk ke dalam paru-paru dan bahkan ke aliran darah, menyebabkan masalah kesehatan seperti gangguan pernapasan, penyakit jantung, dan bahkan kematian.

b. PM_{10}

Particulate Matter 10 (PM_{10}) adalah partikel dalam udara dengan diameter kurang dari 10 mikrometer[14]. Seperti $PM_{2.5}$, partikel ini juga dapat berasal dari berbagai sumber dan dapat mempengaruhi kesehatan manusia ketika terhirup dalam jumlah

yang besar. PM_{10} dapat menyebabkan gangguan pernapasan, iritasi mata, dan masalah kesehatan lainnya.

c. Ozon

Ozon (O_3) adalah gas yang terdiri dari tiga atom oksigen. Ozon di atmosfer berasal dari reaksi kimia antara oksigen, sinar matahari, dan polutan seperti nitrogen oksida dan senyawa organik volatil[14]. Paparan ozon yang tinggi dapat menyebabkan iritasi paru-paru, memperburuk asma, dan meningkatkan risiko penyakit pernapasan lainnya.

d. CO

Karbon Monoksida (CO) adalah gas beracun yang tidak berwarna dan tidak berbau yang dihasilkan dari pembakaran bahan bakar fosil seperti bensin, kayu, atau gas alam[14]. Paparan CO dalam jumlah tinggi dapat menyebabkan keracunan CO, yang dapat mengakibatkan sakit kepala, mual, pusing, bahkan kematian.

e. SO_2

Dioksida Sulfur (SO_2) adalah gas beracun yang dihasilkan dari pembakaran bahan bakar fosil yang mengandung belerang, seperti batu bara dan minyak bumi[14]. Paparan SO_2 dapat menyebabkan iritasi pada saluran pernapasan, masalah pernapasan, dan memperburuk kondisi kesehatan yang ada seperti asma.

f. NO_2

Dioksida Nitrogen (NO_2) adalah gas beracun yang dihasilkan dari pembakaran bahan bakar fosil pada kendaraan bermotor, pembangkit listrik, dan industri[14]. Paparan NO_2 dapat menyebabkan iritasi pada saluran pernapasan, meningkatkan risiko infeksi saluran pernapasan atas, dan memperburuk gejala asma.

2.1.2 Metode Perhitungan Kualitas Udara

Perhitungan kualitas udara umumnya dilakukan menggunakan indeks yang disebut Air Quality Index (AQI). AQI memberikan satu angka yang menggambarkan kualitas udara dan dampaknya terhadap kesehatan manusia.

Langkah-langkah Perhitungan AQI

1. Pengumpulan Data Polutan:

- Data konsentrasi polutan diukur dan dikumpulkan dari stasiun pemantau udara. Data ini biasanya dalam bentuk konsentrasi rata-rata selama periode waktu tertentu (misalnya, per jam, per 24 jam).

2. Konversi ke AQI Sub-Index:

- Setiap polutan memiliki kurva fungsi yang mengkonversi konsentrasi polutan ke AQI sub-index. Rumus dan kurva ini ditentukan oleh badan lingkungan setempat (misalnya, EPA di Amerika Serikat dan ISPU di Indonesia).
- Misalnya, untuk polutan $PM_{2.5}$, AQI sub-index dihitung dengan menginterpolasi nilai konsentrasi $PM_{2.5}$ pada kurva yang telah ditetapkan.

Tabel 1 Tabel Konversi Nilai Konsentrasi Parameter AQI

O ₃ (ppm) 8-hour	O ₃ (ppm) 1-hour ¹	PM ₂₅ (µg/m ³) 24-hour	PM ₁₀ (µg/m ³) 24-hour	CO (ppm) 8-hour	SO ₂ (ppb) 1-hour	NO ₂ (ppb) 1- hour	AQI	Klasifikasi
0.000 - 0.054	-	0.0 – 9.0	0 - 54	0.0 - 4.4	0 - 35	0 - 53	0 - 50	Baik
0.055 - 0.070	-	9.1 – 35.4	55 - 154	4.5 - 9.4	36 - 75	54 - 100	51 - 100	Sedang
0.071 - 0.085	0.125 - 0.164	35.5 – 55.4	155 - 254	9.5 - 12.4	76 - 185	101 - 360	101 - 150	Tidak baik untuk kelompok sensitif
0.086 - 0.105	0.165 - 0.204	(55.5 - 125.4) ³	255 - 354	12.5 - 15.4	³ 186 - 304	361 - 649	151 - 200	Tidak Sehat
0.106 - 0.200	0.205 - 0.404	(125.5 - (225.4) ³	355 - 424	15.5 - 30.4	³ 305 - 604)	650 - 1249	201 - 300	Sangat Tidak Sehat
0.201- (²)	0.405+	225.5+	425+	30.5+	³ 605+	1250+	301+	Hazardous ⁴

¹ Wilayah pada umumnya diwajibkan melaporkan AQI berdasarkan nilai **O₃ rata-rata 8 jam**. Namun, terdapat beberapa wilayah di mana AQI berdasarkan **O₃ rata-**

rata 1 jam dianggap lebih berhati-hati. Dalam kasus ini, selain menghitung nilai indeks **O3 rata-rata 8 jam**, nilai **O3 rata-rata 1 jam** juga dapat dihitung, dan nilai maksimum dari keduanya yang dilaporkan.

² **Nilai O3 rata-rata 8 jam** tidak mendefinisikan AQI tinggi (≥ 301). Nilai AQI sebesar 301 atau lebih dihitung menggunakan konsentrasi **O3 rata-rata 1 jam**.

³ **Konsentrasi SO2 rata-rata 1 jam** tidak mendefinisikan AQI tinggi (≥ 200). Nilai AQI sebesar 200 atau lebih dihitung menggunakan konsentrasi **SO2 rata-rata 24 jam**.

⁴ Nilai AQI di antara batas rentang dihitung menggunakan **persamaan 1** dalam lampiran ini. Untuk nilai AQI dalam kategori berbahaya, nilai AQI lebih dari 500 harus dihitung menggunakan **persamaan 1** dan konsentrasi yang ditentukan untuk nilai AQI 500. Nilai AQI sebesar 500 didefinisikan sebagai berikut[15]:

- **O3 rata-rata 1 jam**: 0.604 ppm
- **PM2.5 rata-rata 24 jam**: 325.4 $\mu\text{g}/\text{m}^3$
- **PM10 rata-rata 24 jam**: 604 $\mu\text{g}/\text{m}^3$
- **CO rata-rata 1 jam**: 50.4 ppm
- **SO2 rata-rata 1 jam**: 1004 ppb
- **NO2 rata-rata 1 jam**: 2049 ppb

- **Persamaan 1**

$$I_p = \frac{I_{Hi} - I_{Lo}}{BP_{Hi} - BP_{Lo}} (C_p - BP_{Lo}) + I_{Lo}.$$

Keterangan :

I_p = Indeks untuk polutan p.

C_p = Konsentrasi polutan p yang telah dipotong/dibatasi.

BP_{Hi} = Titik batas konsentrasi yang lebih besar dari atau sama dengan C_p .

BP_{Lo} = Titik batas konsentrasi yang lebih kecil dari atau sama dengan C_p .

I_{Hi} = Nilai AQI (Indeks Kualitas Udara) yang sesuai dengan BP_{Hi} .

I_{Lo} = Nilai AQI yang sesuai dengan BP_{Lo} .

3. Penentuan AQI Total:

Setelah semua polutan dikonversi ke AQI sub-index masing-masing, nilai tertinggi dari sub-index tersebut diambil sebagai AQI total untuk periode waktu tertentu.

Nilai tertinggi ini menunjukkan polutan yang memiliki dampak paling signifikan pada kualitas udara saat itu.

Misalkan Anda memiliki nilai O₃ rata-rata 8 jam sebesar **0.078 ppm**, nilai PM_{2.5} sebesar **35.9 µg/m³**, dan nilai CO sebesar **8.4 ppm**. Gunakan persamaan tiga kali untuk AQI dengan beberapa parameter polutan:

$$O_3: \frac{150 - 101}{0.085 - 0.071} (0.78 - 0.071) + 101 = 126$$

$$PM_{2.5}: \frac{150 - 101}{55.4 - 35.5} (35.9 - 35.5) + 101 = 102$$

$$CO: \frac{100 - 51}{9.4 - 4.5} (8.4 - 4.5) + 51 = 90$$

AQI tertinggi adalah **126**, dengan O₃ sebagai polutan utama.

4. Kategorisasi AQI:

AQI kemudian dikategorikan ke dalam berbagai level berdasarkan dampaknya terhadap kesehatan. Kategori ini bervariasi antara negara, tetapi umumnya sebagai berikut:

Indeks Kualitas Udara AQI	AQI Level		PM _{2.5} (µg/m ³)	Deskripsi
	0-50	Good	0-12.0	Kualitas udara baik dan bersih tidak memberikan dampak bagi kesehatan
	51-100	Moderate	12.1-35.4	Kualitas udara cukup dapat diterima. Namun, ada risiko bagi beberapa orang, terutama mereka yang sangat sensitif terhadap polusi udara.
	101-150	Unhealthy for sensitive groups	35.5-55.4	Efek kesehatan mungkin dialami oleh sekelompok orang yang sensitif, sedangkan masyarakat umum memiliki risiko yang lebih rendah untuk terpengaruh.
	151-200	Unhealthy	55.5-150.4	sebagian anggota masyarakat umum mungkin mengalami efek kesehatan, sementara anggota kelompok yang sensitif berisiko mengalami efek kesehatan yang lebih serius
	201-300	Very Unhealthy	150.5-250.4	Peringatan kesehatan: Risiko efek kesehatan meningkat untuk semua orang.
	301+	Hazardous	250.5+	Peringatan kesehatan dalam kondisi darurat: semua orang memiliki risiko yang lebih tinggi untuk terpengaruh.

Gambar 1 Kategori Kualitas Udara menurut EPA

AQI bernilai **126** memiliki kategori klasifikasi **tidak sehat untuk kelompok sensitif** dengan deskripsi bahwa efek kesehatan mungkin dialami oleh kelompok orang yang sensitif, sedangkan masyarakat umum memiliki resiko yang lebih rendah untuk terpengaruh.

2.2 Provinsi Lampung

Secara astronomis, Provinsi Lampung terletak di antara $103^{\circ}40'$ hingga $105^{\circ}50'$ Bujur Timur dan $60^{\circ}45'$ sampai $30^{\circ}45'$ Lintang Selatan. Dilihat dari posisi geografisnya, provinsi ini berbatasan dengan Provinsi Sumatera Selatan dan Bengkulu di bagian Utara, Selat Sunda di bagian Selatan, Samudera Indonesia di bagian Barat, serta Laut Jawa di bagian Timur. Provinsi Lampung terbagi menjadi 13 kabupaten dan 2 kota, meliputi Kabupaten Lampung Barat, Tanggamus, Lampung Selatan, Lampung Timur, Lampung Tengah, Lampung Utara, Way Kanan, Tulang Bawang, Pesawaran, Pringsewu, Mesuji, Tulang Bawang Barat, dan Pesisir Barat, serta Kota Bandar Lampung dan Metro[16].

2.3 Peramalan (*Forecasting*)

Forecasting adalah proses perkiraan atau prediksi tentang peristiwa masa depan berdasarkan informasi dan data yang tersedia saat ini. Ini melibatkan penggunaan metode statistik, analisis data, dan model matematika untuk membuat perkiraan tentang apa yang mungkin terjadi di masa mendatang. Tujuan dari forecasting adalah untuk memberikan pandangan yang lebih baik tentang kemungkinan peristiwa di masa depan, sehingga organisasi atau individu dapat membuat keputusan yang lebih baik dan merencanakan tindakan yang sesuai [17]. Metode forecasting dapat digunakan dalam berbagai bidang, termasuk ekonomi, keuangan, bisnis, cuaca, dan lainnya.

2.4 *Sequential Data* (Data Berurutan)

Data sequential adalah jenis data di mana urutan elemen sangat penting dan mempengaruhi analisis serta pengolahan data tersebut. Dalam data sequential, setiap elemen data memiliki hubungan dengan elemen sebelumnya dan berikutnya, sehingga perubahan atau tindakan pada satu elemen dapat mempengaruhi elemen-elemen lainnya dalam urutan tersebut. Contoh data sequential termasuk teks dalam

kalimat, sinyal suara, data time-series seperti harga saham atau cuaca, dan tindakan dalam permainan catur.[13]

Dalam analisis dan pemodelan data sequential, pendekatan khusus seperti model time-series, jaringan saraf berulang (RNN), Long Short-Term Memory (LSTM), dan Gated *Recurrent* Unit (GRU) sering digunakan. RNN adalah model dasar untuk data sequential, tetapi memiliki keterbatasan dalam mengingat informasi jangka panjang. LSTM dan GRU adalah varian dari RNN yang dirancang untuk mengatasi masalah ini dengan mekanisme khusus yang membantu model untuk menyimpan informasi lebih lama dan menghindari masalah vanishing gradient.

2.5 Korelasi Pearson

Korelasi Pearson mengukur kekuatan dan arah hubungan linear antara dua variabel. Nilai korelasi Pearson berkisar dari -1 hingga 1, di mana:

- 1 menunjukkan hubungan linear positif yang sempurna.
- -1 menunjukkan hubungan linear negatif yang sempurna.
- 0 menunjukkan tidak ada hubungan linear antara variabel.

Analisis koefisien korelasi digunakan untuk menentukan arah dan kekuatan hubungan antara dua atau lebih variabel. Arah hubungan bisa berupa positif atau negatif, sementara kekuatan hubungan ditunjukkan oleh besarnya koefisien korelasi. (Sugiyono, 2017).

Berdasarkan pedoman interpretasi dan analisis koefisien korelasi menurut Sugiyono:

- - 0,00 – 0,199: sangat rendah
- - 0,20 – 0,399: rendah
- - 0,40 – 0,599: sedang
- - 0,60 – 0,799: kuat
- - 0,80 – 1,000: sangat kuat

2.6 Artificial Intelligence (AI)

Secara historis, para peneliti telah mengejar beberapa versi AI yang berbeda. Beberapa mendefinisikan kecerdasan berdasarkan kesesuaian dengan kinerja manusia, sementara yang lain lebih memilih definisi formal abstrak dari kecerdasan yang disebut rasionalitas—secara longgar, melakukan hal yang "benar." Subjeknya sendiri juga bervariasi: beberapa menganggap kecerdasan sebagai sifat dari proses pemikiran dan penalaran internal, sementara yang lain berfokus pada perilaku cerdas, suatu karakterisasi eksternal.

Dari dua dimensi ini—manusia vs. rasional dan pemikiran vs. perilaku—terdapat empat kombinasi yang mungkin, dan ada pendukung serta program penelitian untuk keempatnya. Metode yang digunakan tentunya berbeda: pengejaran kecerdasan seperti manusia harus sebagian berupa ilmu empiris yang terkait dengan psikologi, melibatkan pengamatan dan hipotesis tentang perilaku dan proses pemikiran manusia yang sebenarnya; pendekatan rasionalis, di sisi lain, melibatkan kombinasi matematika dan teknik, serta berhubungan dengan statistik, teori kendali, dan ekonomi[13]

AI (*Artificial Intelligence*) adalah bidang ilmu komputer yang bertujuan untuk membuat sistem komputer yang mampu melakukan tugas-tugas yang biasanya memerlukan kecerdasan manusia. Ini mencakup pembuatan algoritma dan model matematika yang memungkinkan komputer untuk belajar dari data, memahami pola-pola kompleks, membuat keputusan, dan menyelesaikan tugas-tugas dengan sedikit atau tanpa campur tangan manusia. AI mencakup berbagai teknik seperti *machine learning*, *deep learning*, dan lainnya.

Big data telah secara bertahap masuk ke semua aspek kehidupan. Sumber daya data menjadi kekayaan kritis di masa depan. Penerapan pemikiran big data dan teknologi diagnosis kecerdasan buatan (AI) dalam tata kelola lingkungan dapat memberikan dukungan data dan teknis untuk tata kelola lingkungan publik. Selain itu, tata kelola lingkungan dapat memberikan gagasan ilmiah dan akurat untuk pengambilan keputusan pemerintah dalam pemantauan lingkungan publik dan peringatan dini melalui pengumpulan data, pemantauan real-time, dan manajemen partisipasi warga[18].

2.7 Machine Learning

Seorang agen dikatakan *learning* jika ia meningkatkan kinerjanya setelah melakukan observasi tentang dunia. *Learning* bisa berkisar dari hal yang sepele, seperti mencatat daftar belanja, hingga hal yang mendalam, seperti ketika Albert Einstein menyimpulkan teori baru tentang alam semesta. Ketika agen tersebut adalah komputer, kita menyebutnya *machine learning*: komputer mengamati sejumlah data, membangun model berdasarkan data tersebut, dan menggunakan model ini sebagai hipotesis tentang dunia sekaligus sebagai perangkat lunak yang dapat memecahkan masalah.

Machine learning memiliki beberapa metode utama untuk melatih model dan membuat prediksi berdasarkan data. *Supervised learning* menggunakan dataset yang diberi label untuk memprediksi *output* yang benar, sementara *unsupervised learning* menemukan pola dalam data tanpa label. *Semi-supervised learning* mengombinasikan keduanya dengan menggunakan sebagian kecil data yang diberi label. *Reinforcement learning* melibatkan agen yang belajar melalui umpan balik dari lingkungan untuk memaksimalkan reward. *Deep learning*, sebagai subset dari *machine learning*, menggunakan jaringan saraf tiruan dengan banyak lapisan untuk menganalisis dan belajar dari data, efektif dalam pengolahan citra, pengenalan suara, dan pemrosesan bahasa alami[13].

2.8 Deep learning

Deep learning adalah keluarga teknik yang luas dalam *machine learning* di mana hipotesis diambil dalam bentuk sirkuit aljabar kompleks dengan kekuatan koneksi yang dapat disesuaikan. Kata "*deep*" mengacu pada fakta bahwa sirkuit tersebut biasanya diorganisir menjadi banyak lapisan, yang berarti bahwa jalur komputasi dari input ke *output* memiliki banyak langkah. *Deep learning* saat ini adalah pendekatan yang paling banyak digunakan untuk aplikasi seperti visual object recognition, *machine translation*, speech recognition, speech synthesis, dan image synthesis; juga berperan penting dalam aplikasi reinforcement *learning*.

Deep learning memiliki asal usul dalam pekerjaan awal yang mencoba memodelkan jaringan *neuron* di otak (McCulloch dan Pitts, 1943) dengan sirkuit komputasional. Karena alasan ini, jaringan yang dilatih dengan metode *deep*

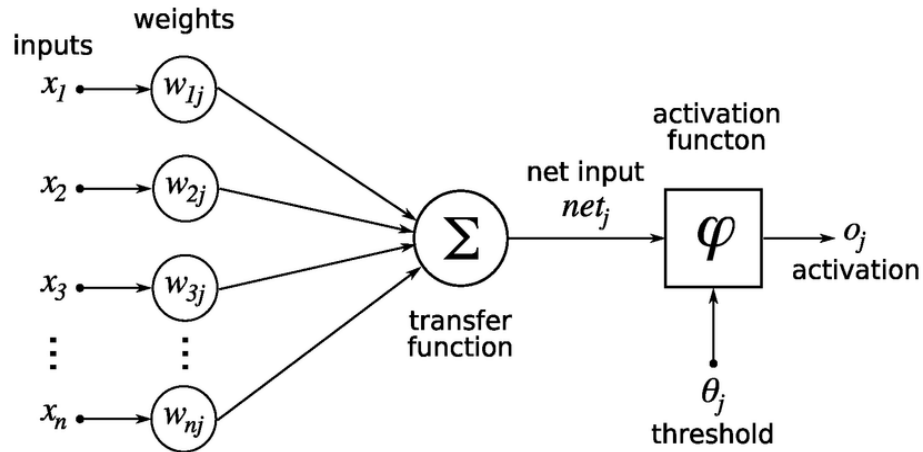
learning sering disebut sebagai *neural networks*, meskipun kemiripannya dengan sel dan struktur saraf yang sebenarnya sangatlah dangkal.[13]

2.8.1 Jaringan Syaraf Tiruan (JST)

Arsitektur atau struktur Jaringan Syaraf Tiruan (JST) atau *Artificial Neural Network* (ANN) menggambarkan susunan lapisan dan *neuron* pada input, *hidden*, dan *output* yang terhubung dengan bobot, fungsi aktivasi, dan fungsi pembelajaran (***learning function***). Cara neuron dari sebuah neural network disusun memiliki hubungan erat dengan learning algorithm yang digunakan untuk melatih network tersebut. Oleh karena itu, kita dapat berbicara tentang learning algorithms (rules) yang digunakan dalam desain neural networks sebagai sesuatu yang terstruktur. Menurut Haykin, ada tiga kelas dasar arsitektur ANN: jaringan dengan satu lapisan umpan maju (***Single Layer Feedforward Network***), jaringan dengan *multi* lapisan umpan maju (***Multilayer Feedforward Network***), dan jaringan berulang (***Recurrent Network***)[19].

1. *Single Layer Feedforward Network*

Arsitektur jaringan ini terdiri dari dua lapisan: lapisan input dan lapisan *output*. Lapisan input menerima sinyal data, sementara lapisan *output* memberikan hasil. *Neuron-neuron* di lapisan input terhubung dengan bobot ke lapisan *output* dalam satu arah maju. Meskipun ada dua lapisan, ini disebut jaringan satu lapisan karena hanya lapisan *output* yang melakukan komputasi tanpa lapisan lain di antaranya seperti yang terlihat pada gambar 2.

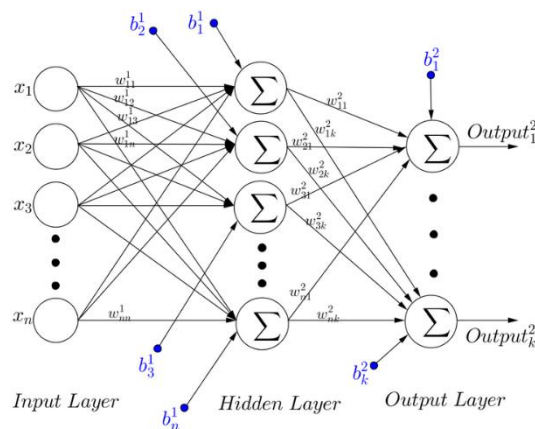


Gambar 2 ANN Satu Lapisan Umpan Maju (*Single Layer feedforward network*)

2. Multilayer Feedforward Network

Jaringan ini memiliki lapisan tambahan antara lapisan input dan *output* yang disebut lapisan *hidden*. Lapisan *hidden* ini melakukan perhitungan dari input yang kemudian diteruskan ke *output*. Jaringan *multilayer* dapat memiliki lebih dari satu lapisan *hidden*, tergantung pada kompleksitas masalah yang dihadapi[19].

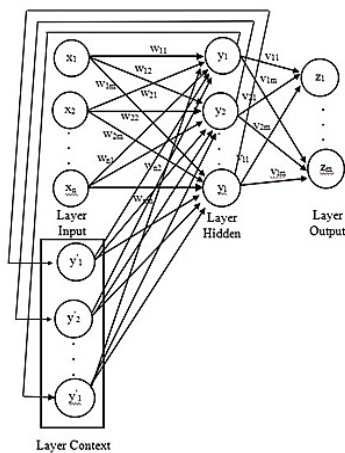
Lapisan *hidden* pada jaringan ini menerima input dari lapisan input melalui bobot (w) dan bias (b), yang kemudian diproses menggunakan fungsi aktivasi untuk memperkenalkan non-linearitas. Output dari lapisan *hidden* diteruskan ke lapisan *output*, di mana bobot dan bias tambahan digunakan untuk menghasilkan prediksi akhir seperti pada gambar 3.



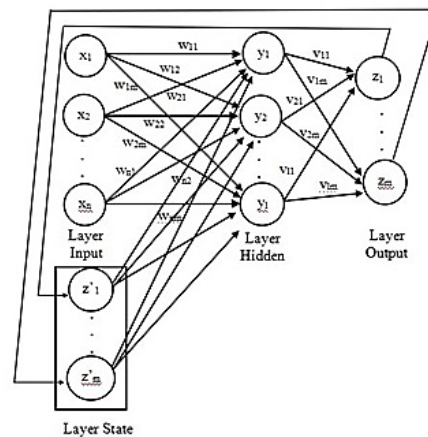
Gambar 3 ANN Multilayer Umpan Maju (*Multilayer feedforward network*)

3. Recurrent Network

Berbeda dengan arsitektur umpan maju, jaringan berulang (*Recurrent Network*) memiliki minimal satu umpan balik (*loop*) untuk meningkatkan kemampuannya mempelajari karakter sementara dari data yang diberikan[19]. Beberapa jenis jaringan berulang telah dikembangkan, termasuk model Elman (gambar 4) dan Jordan (gambar 5). Model Elman [32] membuat salinan *neuron* lapisan *hidden* pada lapisan input yang disebut *context input*, berfungsi menyimpan status sebelumnya dari lapisan *hidden* dan menyampaikannya kembali. Konteks input ini terhubung penuh dengan lapisan *hidden* dan memiliki bobot 1. Model Jordan [33] membuat salinan lapisan *output* pada lapisan input yang disebut *state layer*. Dengan proses ini, hasil *output* dari iterasi sebelumnya menjadi bagian dari input pada iterasi berikutnya.



Gambar 4 Arsitektur ANN berulang (*Recurrent Network*) model Elman

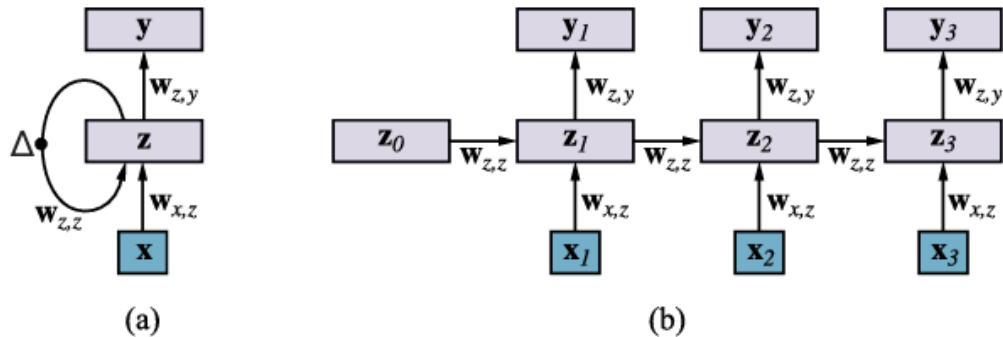


Gambar 5 Arsitektur ANN berulang (*Recurrent Network*) model Jordan

2.8.2 Single Recurrent Neural Network (RNN)

Recurrent Neural Networks (RNNs) memiliki keunikan dalam memproses data sekuensial dengan mengingat informasi dari masa lalu. Berbeda dengan *Artificial Neural Networks* (ANNs) biasa, RNN tidak mengabaikan informasi dari masa lalu dalam proses pembelajarannya. RNN memanfaatkan *looping* dalam arsitekturnya untuk menyimpan informasi tersebut. Secara umum, RNN adalah bagian dari keluarga *Neural Network* yang dirancang khusus untuk memproses data sekuensial.

Cara kerjanya adalah dengan melakukan *looping*, memungkinkan informasi untuk berpindah dari satu langkah ke langkah berikutnya dalam jaringan.



Gambar 6 Arsitektur RNN

Gambar 6(a) menunjukkan RNN dasar di mana lapisan tersembunyi (z) memiliki koneksi berulang yang memungkinkan jaringan menyimpan informasi dari langkah waktu sebelumnya melalui hidden state. Simbol Δ menunjukkan adanya penundaan atau propagasi informasi antar langkah waktu. Pada setiap time step (t), input (x_t) dikombinasikan dengan hidden state sebelumnya (z_{t-1}), yang kemudian diproses untuk memperbarui hidden state saat ini (z_t). Proses ini menggunakan bobot ($W_{x,z}$) untuk input dan ($W_{z,z}$) untuk koneksi rekuren pada hidden state. Selanjutnya, hidden state (z_t) digunakan untuk menghasilkan output (y_t) melalui bobot ($W_{z,y}$) [13].

Pada gambar 6(b), jaringan RNN "dibuka" atau *unrolled* untuk beberapa langkah waktu, yang menggambarkan bagaimana perhitungan dilakukan secara berurutan dari ($t=1$) hingga ($t=3$). Dalam bentuk ini, RNN terlihat seperti jaringan feedforward di mana bobot ($W_{x,z}$, $W_{z,z}$) dan ($W_{z,y}$) tetap sama di semua langkah waktu. Model ini memungkinkan RNN untuk memproses data sekuensial dengan mengakumulasi informasi dari langkah waktu sebelumnya, sehingga dapat menangkap pola temporal dalam urutan data. Output (y_t) dan input (x_t) diamati pada setiap langkah waktu, yang digunakan selama proses pelatihan untuk meminimalkan kesalahan prediksi dan mengoptimalkan bobot jaringan [13].

Sehingga Proses perhitungan Single RNN dapat dijabarkan sebagai berikut.

1. Input Data

Setiap data dalam urutan diberikan ke RNN satu per satu pada setiap time step. Misalnya, jika kita memiliki urutan data X_1, X_2, \dots, X_T , data ini diberikan ke RNN secara berurutan.

2. Hidden Layer

RNN memiliki state tersembunyi (z_t) yang diperbarui pada setiap time step berdasarkan input saat ini (x_t) dan state tersembunyi sebelumnya (z_{t-1}). State tersembunyi ini berfungsi sebagai memori internal yang menyimpan informasi tentang urutan data sebelumnya.

3. Proses Perhitungan

Pada setiap time step t , RNN menghitung state tersembunyi baru (h_t) menggunakan formula berikut:

$$z_t = \sigma (W_{z,z} z_{t-1} + W_{x,z} x_t + b_z)$$

di mana, z_t adalah hidden state RNN, z_{t-1} adalah hidden state sebelumnya, $W_{z,z}$ adalah bobot untuk state tersembunyi sebelumnya, $W_{x,z}$ adalah bobot untuk input saat ini. x_t adalah input, b_z adalah bias, σ adalah fungsi aktivasi sigmoid[13]

4. Output

RNN dapat menghasilkan *output* pada setiap time step atau hanya pada time step terakhir, tergantung pada tugas yang dihadapi. *Output* pada setiap time step (y_t) dapat dihitung menggunakan formula berikut:

$$y_t = \phi (W_{z,y} z_t + b_y)$$

di mana, $W_{z,y}$ adalah bobot untuk state tersembunyi ke output, b_y adalah bias, ϕ adalah fungsi aktivasi (misalnya fungsi softmax)[13].

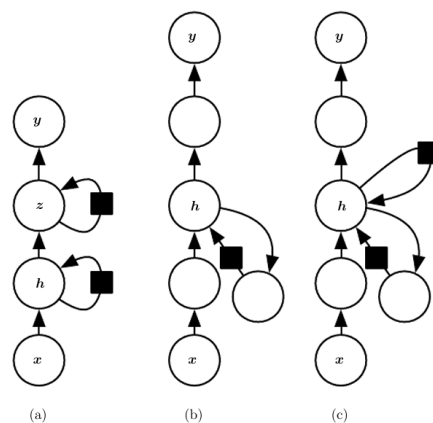
5. Kelemahan dan Masalah RNN

Meskipun RNN efektif dalam memodelkan data sekuensial, mereka memiliki beberapa kelemahan:

- **Vanishing Gradient Problem:** Gradien dapat menjadi sangat kecil selama backpropagation, menyebabkan jaringan sulit untuk dilatih untuk urutan yang panjang.
- **Exploding Gradient Problem:** Gradien dapat menjadi sangat besar, menyebabkan instabilitas dalam pelatihan.

6. Deep Recurrent Neural Network

Komputasi dalam sebagian besar RNN dapat diuraikan menjadi tiga blok parameter: transformasi dari input ke keadaan tersembunyi, dari keadaan tersembunyi sebelumnya ke yang berikutnya, dan dari keadaan tersembunyi ke output. Arsitektur RNN yang diilustrasikan pada gambar 10 menggunakan satu matriks bobot untuk masing-masing blok ini, sehingga ketika jaringan diurai, tiap blok berperan sebagai transformasi dangkal, yakni transformasi dengan satu lapisan dalam MLP yang dalam, terdiri dari transformasi afine yang dipelajari diikuti oleh nonlinieritas tetap. Memperkenalkan kedalaman tambahan dalam operasi ini dinilai bermanfaat berdasarkan bukti eksperimen (Graves et al., 2013; Pascanu et al., 2014a), yang menunjukkan bahwa kedalaman cukup diperlukan untuk memenuhi kompleksitas pemetaan yang dibutuhkan. Beberapa penelitian awal (Schmidhuber 1992; El Hiji dan Bengio 1996; Graves et al., 2013; Jaeger 2007a) juga mendukung bahwa pemecahan keadaan tersembunyi menjadi lapisan-lapisan hierarkis membawa manfaat, seperti diilustrasikan pada gambar 10. Pascanu et al. (2014a) mengusulkan adanya MLP terpisah untuk tiap blok, yang meskipun meningkatkan kapasitas representasi, dapat mempersulit optimasi jaringan akibat jalur terpendek antara dua variabel waktu yang menjadi lebih panjang, sehingga sulit mengoptimalkan jaringan yang lebih dalam ini.[20]



Gambar 7 Sebuah *recurrent neural network* dapat dibuat lebih dalam dengan berbagai cara (Pascanu et al., 2014a).

- Gambar 7(a) Keadaan *recurrent* tersembunyi dapat dipecah menjadi kelompok-kelompok yang diorganisasikan secara hierarkis[20].
- Gambar 7(b) Komputasi yang lebih dalam (misalnya, MLP) dapat diperkenalkan pada bagian input-ke-tersembunyi, tersembunyi-ke-tersembunyi, dan tersembunyi-ke-output. Ini dapat memperpanjang jalur terpendek yang menghubungkan langkah-langkah waktu yang berbeda[20].
- Gambar 7(c) Efek pemanjangan jalur ini dapat dikurangi dengan memperkenalkan koneksi loncatan pada jalur tersembunyi-ke-tersembunyi, seperti diilustrasikan pada gambar 10c[20].

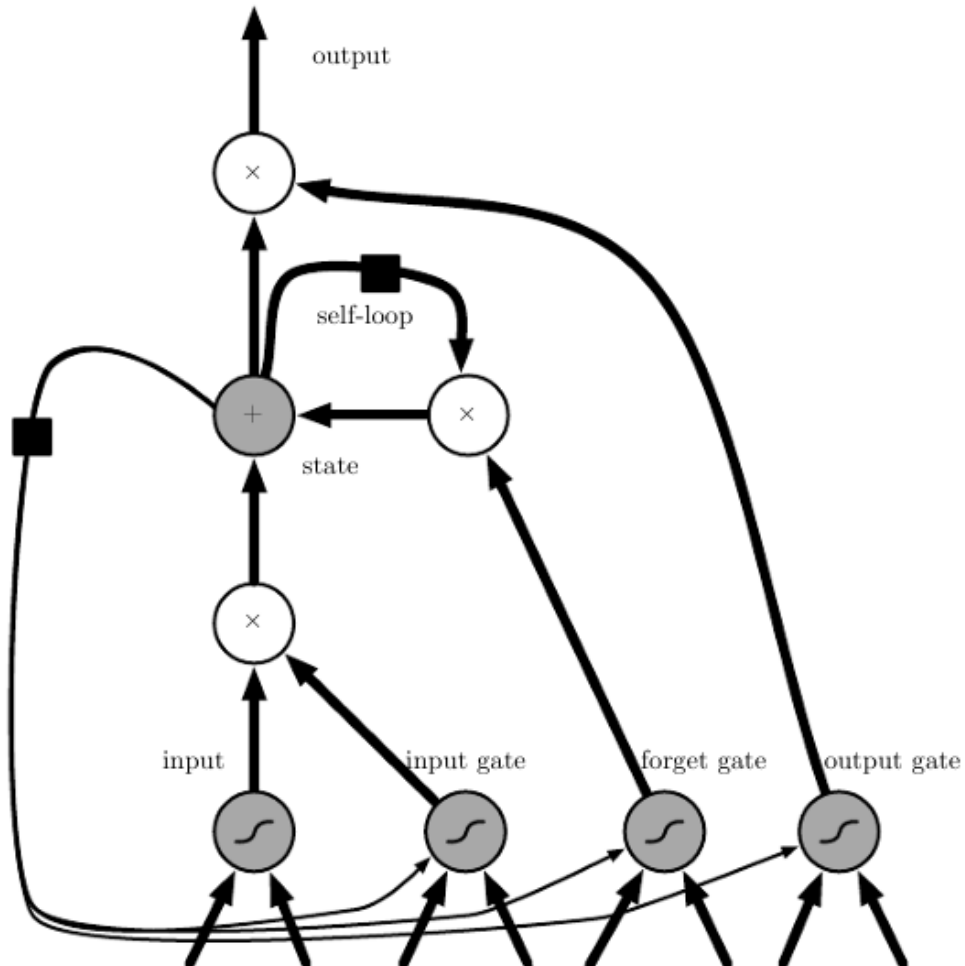
2.8.3 Long short-term memory (LSTM)

Long Short-Term Memory (LSTM) adalah jenis arsitektur jaringan saraf tiruan yang dirancang khusus untuk memproses dan memodelkan data berurutan, seperti teks, audio, atau rangkaian waktu. LSTM menggunakan mekanisme gerbang yang memungkinkannya untuk mengontrol aliran informasi dalam jaringan, yang terdiri dari tiga jenis gerbang: input gate, forget gate, dan output gate[13]. Gerbang-gerbang ini memungkinkan LSTM untuk memilih informasi mana yang disimpan, dilupakan, dan dilewati dari satu langkah waktu ke langkah waktu berikutnya, sehingga menjaga aliran informasi dalam jangka waktu yang panjang[20].

Gagasan cerdas untuk memperkenalkan **self-loops** guna menghasilkan jalur di mana gradien dapat mengalir dalam durasi panjang adalah kontribusi utama dari model **long short-term memory (LSTM)** awal (Hochreiter dan Schmidhuber, 1997). Penambahan penting lainnya adalah menjadikan bobot pada self-loop ini bergantung pada konteks, bukan tetap (Gers et al., 2000). Dengan membuat bobot self-loop ini *gated* (dikendalikan oleh unit tersembunyi lainnya), skala waktu integrasi dapat diubah secara dinamis[20].

Dalam konteks ini, bahkan untuk LSTM dengan parameter tetap, skala waktu integrasi dapat berubah berdasarkan urutan input, karena konstanta waktu dihasilkan oleh model itu sendiri. LSTM telah terbukti sangat berhasil dalam berbagai aplikasi, seperti pengenalan tulisan tangan tak terbatas (Graves et al., 2009), pengenalan suara (Graves et al., 2013; Graves dan Jaitly, 2014), generasi tulisan tangan (Graves, 2013), penerjemahan mesin (Sutskever et al., 2014),

pembuatan caption gambar (Kiros et al., 2014b; Vinyals et al., 2014b; Xu et al., 2015), dan *parsing* (Vinyals et al., 2014a)[20].



Gambar 8 LSTM

Diagram blok LSTM diilustrasikan dalam Gambar Persamaan *forward propagation* yang sesuai diberikan di bawah ini untuk kasus arsitektur jaringan rekursif dangkal. Arsitektur yang lebih dalam juga telah berhasil digunakan (Graves et al., 2013; Pascanu et al., 2014a)[20].

Alih-alih sebuah unit yang hanya menerapkan *non-linearity* secara elemen pada transformasi affine dari input dan unit rekursif, jaringan rekursif LSTM memiliki "sel LSTM" yang memiliki rekursi internal (*self-loop*), selain rekursi eksternal dari RNN. Setiap sel memiliki input dan output yang sama seperti jaringan rekursif

biasa, tetapi dengan lebih banyak parameter dan sistem unit gerbang yang mengontrol aliran informasi[20].

Komponen terpenting adalah unit keadaan s_i^t , yang memiliki self-loop linear serupa dengan *leaky units* yang dijelaskan di bagian sebelumnya. Namun, di sini, bobot self-loop (atau konstanta waktu terkait) dikontrol oleh unit gerbang pelupa f_i^t untuk langkah waktu t dan sel i , yang menetapkan bobot ini ke nilai antara 0 dan 1 melalui unit sigmoid:

$$f_i^{(t)} = \sigma \left(b_i^f + \sum_j U_{i,j}^f x_j^{(t)} + \sum_j W_{i,j}^f h_j^{(t-1)} \right),$$

di mana $\mathbf{x}^{(t)}$ adalah vektor input saat ini dan $\mathbf{h}^{(t)}$ adalah vektor lapisan tersembunyi saat ini yang berisi output dari semua sel LSTM, serta \mathbf{b}^f , \mathbf{U}^f , dan \mathbf{W}^f masing-masing adalah bias, bobot input, dan bobot rekursif untuk gerbang pelupa. Keadaan internal sel LSTM diperbarui sebagai berikut, tetapi dengan bobot self-loop kondisional $f_i^{(t)}$:

$$s_i^{(t)} = f_i^{(t)} s_i^{(t-1)} + g_i^{(t)} \sigma \left(b_i + \sum_j U_{i,j} x_j^{(t)} + \sum_j W_{i,j} h_j^{(t-1)} \right),$$

di mana \mathbf{b} , \mathbf{U} , dan \mathbf{W} masing-masing menunjukkan bias, bobot input, dan bobot rekursif dalam sel LSTM. Unit gerbang input eksternal $g_i^{(t)}$ dihitung serupa dengan gerbang pelupa (dengan unit sigmoid untuk mendapatkan nilai antara 0 dan 1), tetapi dengan parameter tersendiri:

$$g_i^{(t)} = \sigma \left(b_i^g + \sum_j U_{i,j}^g x_j^{(t)} + \sum_j W_{i,j}^g h_j^{(t-1)} \right).$$

Output $h_i^{(t)}$ dari sel LSTM juga dapat dimatikan melalui gerbang output $q_i^{(t)}$, yang juga menggunakan unit sigmoid untuk *gating*:

$$h_i^{(t)} = \tanh(s_i^{(t)}) q_i^{(t)},$$

$$q_i^{(t)} = \sigma \left(b_i^o + \sum_j U_{i,j}^o x_j^{(t)} + \sum_j W_{i,j}^o h_j^{(t-1)} \right),$$

yang memiliki parameter \mathbf{b}^o , \mathbf{U}^o , dan \mathbf{W}^o untuk bias, bobot input, dan bobot rekursifnya. Di antara variasi LSTM, seseorang dapat memilih untuk menggunakan keadaan sel $s_i^{(t)}$ sebagai input tambahan (dengan bobot tersendiri) ke dalam tiga gerbang unit ke-ii, seperti yang ditunjukkan pada Gambar 8. Hal ini memerlukan tiga parameter tambahan[20].

1. Kelebihan dan Kekurangan long short term memory (LSTM)

a. Kelebihan:

1. **Kemampuan Menangani Data Sekuensial:** Efektif dalam mempelajari ketergantungan jangka panjang dalam data sekuensial.
2. **Mengatasi Vanishing Gradient:** Struktur gating mencegah gradien menghilang atau meledak selama pelatihan.
3. **Performa Tinggi:** Digunakan secara luas di berbagai aplikasi seperti *speech recognition* dan *natural language processing*.

b. Kekurangan:

1. **Komputasi Berat:** Membutuhkan waktu pelatihan lebih lama dan sumber daya komputasi lebih besar.
2. **Kebutuhan Data Besar:** Membutuhkan banyak data untuk mendapatkan hasil optimal.
3. **Sulit Diinterpretasi:** Kompleksitas arsitektur membuat model sulit dipahami secara intuitif.

Secara keseluruhan, LSTM adalah alat yang ampuh untuk prediksi kualitas udara. Meskipun memiliki beberapa kekurangan, seperti kebutuhan data yang banyak dan komputasi yang intensif, kelebihanannya dalam menangani data sekuensial, memori jangka panjang, dan performa yang akurat menjadikannya pilihan yang menarik untuk penelitian prediksi kualitas udara.

2.8.4 Gated Recurrent Unit (GRU)

Gated Recurrent Unit (GRU) adalah jenis arsitektur jaringan saraf rekuren (RNN) yang dikembangkan untuk mengatasi beberapa masalah yang dimiliki oleh RNN standar, seperti masalah vanishing gradient dan kemampuan untuk menangkap dependensi temporal dalam urutan data panjang[20].

Perbedaan utama dengan LSTM adalah bahwa satu unit gerbang secara bersamaan mengontrol faktor pelupaan dan keputusan untuk memperbarui unit state. Persamaan pembaruannya adalah sebagai berikut:

$$\mathbf{h}_i^{(t)} = \mathbf{u}_i^{(t-1)} \mathbf{h}_i^{(t-1)} + (\mathbf{1} - \mathbf{u}_i^{(t-1)}) \sigma \left(\mathbf{b}_i + \sum_j \mathbf{U}_{ij} \mathbf{x}_j^{(t-1)} + \sum_j \mathbf{W}_{ij} \mathbf{r}_j^{(t-1)} \mathbf{h}_j^{(t-1)} \right),$$

di mana \mathbf{u} adalah gerbang "update" dan \mathbf{r} adalah gerbang "reset". Nilai mereka didefinisikan sebagai:

$$\mathbf{u}_i^{(t)} = \sigma \left(\mathbf{b}_i^{(u)} + \sum_j \mathbf{U}_{ij}^u \mathbf{x}_j^{(t)} + \sum_j \mathbf{W}_{ij}^u \mathbf{h}_j^{(t)} \right),$$

dan

$$\mathbf{r}_i^{(t)} = \sigma \left(\mathbf{b}_i^{(r)} + \sum_j \mathbf{U}_{ij}^r \mathbf{x}_j^{(t)} + \sum_j \mathbf{W}_{ij}^r \mathbf{h}_j^{(t)} \right).$$

Gerbang reset dan update dapat secara individual "mengabaikan" bagian-bagian dari vektor state. Gerbang update bertindak seperti *leaky integrator* bersyarat yang dapat mengontrol setiap dimensi secara linear, sehingga dapat memilih untuk menyalinnya (pada salah satu ujung sigmoid) atau mengabaikannya sepenuhnya (pada ujung lainnya) dengan menggantikannya dengan nilai "state target" baru (menuju konvergensi yang diinginkan oleh leaky integrator)[20].

Gerbang reset mengontrol bagian dari state mana yang akan digunakan untuk menghitung state target berikutnya, memperkenalkan efek non-linear tambahan dalam hubungan antara state masa lalu dan state masa depan[20].

2. GRU memiliki beberapa keunggulan dibandingkan RNN tradisional:
 1. **Faster Training:** GRU lebih cepat dalam melatih model karena memiliki lebih sedikit parameter.
 2. **Simpler Architecture:** GRU memiliki struktur yang lebih sederhana dibandingkan LSTM, yang membuatnya lebih mudah diimplementasikan dan diinterpretasikan.

2.9 Fitur *Dropout* dan Teknik *Bidirectional*

Dropout dan teknik *Bidirectional* adalah dua konsep penting dalam *deep learning* yang membantu meningkatkan performa dan generalisasi model. Berikut adalah penjelasan mengenai cara kerja masing-masing:

2.9.1 *Dropout*

Dropout adalah teknik regularisasi yang digunakan untuk mencegah overfitting dalam jaringan *neural*. [21] Cara kerjanya sebagai berikut:

1. **Selama Training:**

Pada setiap iterasi, unit (*neurons*) di *layer* tersembunyi dipilih secara acak untuk dinonaktifkan (*dropped out*) dengan probabilitas p . *Neurons* yang dipilih untuk *Dropout* tidak berkontribusi pada *forward pass* dan *backward pass*. *Dropout* hanya diterapkan selama fase pelatihan, bukan pada fase inferensi (*testing*) [21].

2. **Selama Testing/Inferensi:**

Semua *neurons* diaktifkan, tetapi *output* dari setiap *neuron* diskalakan dengan faktor $1-p$ untuk mempertahankan ekspektasi *output* yang sama seperti saat training. di mana p adalah probabilitas dropout yang digunakan selama training [21].

3. Keuntungan *Dropout*:

Membantu dalam mengurangi overfitting dengan membuat jaringan *neural* menjadi lebih robust dan mengurangi ketergantungan pada *neurons* individu dan memaksa jaringan untuk belajar representasi yang lebih umum dan tidak bergantung pada fitur tertentu[21].

2.9.2 Teknik Bidirectional

Bidirectional Neural Networks, khususnya *Bidirectional Recurrent Neural Networks* (Bi-RNNs) dan *Bidirectional Long Short-Term Memory* (Bi-LSTMs) serta *Bidirectional Gated Recurrent Unit* (Bi-GRU) adalah jenis jaringan *neural* yang dapat memiliki informasi dari masa lalu dan masa depan dalam urutan data[22]. Cara kerjanya sebagai berikut:

1. Arsitektur:

Terdiri dari dua lapisan RNN/LSTM/GRU yang berjalan paralel. Satu lapisan memproses input dari awal hingga akhir (*forward direction*). Lapisan lainnya memproses input dari akhir hingga awal (*backward direction*)[22].

2. *Forward Pass*:

Lapisan *forward* dan *backward* masing-masing menghasilkan serangkaian *output*. *Output* dari kedua lapisan ini digabungkan (biasanya dengan concatenation atau penjumlahan) untuk menghasilkan *output* final di setiap waktu[22].

3. Keuntungan *Bidirectional*:

Menggunakan konteks dari kedua arah (masa lalu dan masa depan) untuk membuat keputusan, yang sangat berguna dalam tugas-tugas seperti pemrosesan bahasa alami di mana konteks dari seluruh kalimat penting untuk memahami arti dari setiap kata. Serta Meningkatkan akurasi model dalam tugas-tugas urutan (sequence tasks) seperti speech recognition, tagging, translation dan prediksi[22].

2.10 Fungsi Aktivasi

Fungsi aktivasi memainkan peran penting dalam jaringan saraf tiruan, mengubah jumlah bobot dari input menjadi sinyal *output*. Ini menentukan *output* dari jaringan saraf, biasanya untuk mengklasifikasikan sebagai "ya" atau "tidak." Fungsi aktivasi memetakan nilai hasil dalam rentang tertentu, seperti 0 hingga 1 atau -1 hingga 1 (tergantung pada fungsi yang digunakan).

Fungsi aktivasi umumnya dibagi menjadi dua jenis:

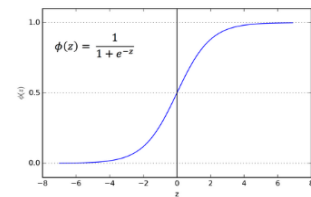
1. Fungsi Aktivasi Linier (Identitas)

Fungsinya linear, tanpa rentang batas ($-\infty$ hingga ∞), sehingga kurang optimal untuk menangani keragaman data dalam jaringan saraf.

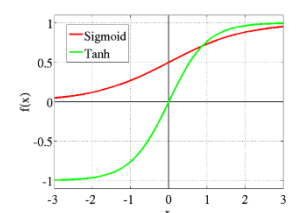
2. Fungsi Aktivasi Non-Linier

Non-linieritas pada fungsi ini membantu model untuk generalisasi data dan membedakan *output*.

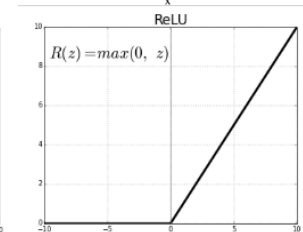
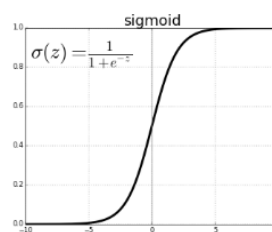
1. **Sigmoid:** Bentuk S, rentang 0 hingga 1, cocok untuk prediksi probabilitas. Sifat diferensialnya memungkinkan perhitungan kemiringan, namun dapat membuat jaringan "stuck" saat pelatihan[19].



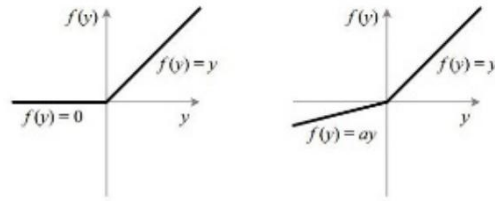
2. **Tanh:** Bentuk S dengan rentang -1 hingga 1, memetakan input negatif dan nol lebih baik. Sering digunakan untuk klasifikasi biner[19].



3. **ReLU:** Fungsi aktivasi yang paling umum saat ini. *Outputnya* nol jika input negatif dan setara dengan input jika positif (0 hingga ∞). Namun, input negatif langsung menjadi nol, mengurangi kemampuan model[19].



4. **Leaky ReLU:** Memperbaiki kelemahan ReLU dengan membiarkan sedikit nilai negatif lewat ($a = 0,01$), menjaga *output* pada rentang $(-\infty$ hingga $\infty)$ [19].



2.11 Index Performa

Indeks performa adalah metrik atau ukuran yang digunakan untuk mengevaluasi kinerja atau efektivitas suatu sistem, model, atau proses. Dalam konteks *machine learning* atau statistik, indeks performa sering digunakan untuk mengevaluasi kualitas prediksi atau model, baik dalam klasifikasi maupun regresi[19].

Dalam konteks regresi, kesalahan model adalah perbedaan antara titik data dan garis tren yang dihasilkan oleh algoritma. Karena ada beberapa titik data, kesalahan dapat ditentukan dengan berbagai cara.

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

Di mana:

- n = jumlah data
- y_i = nilai aktual ke - i
- \hat{y}_i = nilai prediksi ke - i

Mean Absolute Error adalah mean dari nilai absolut kesalahan. Ini adalah metrik yang paling mudah dipahami, karena ini hanya kesalahan rata-rata.

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

Mean Squared Error adalah mean dari kesalahan kuadrat. Ini lebih populer daripada Mean Absolute Error karena fokusnya lebih diarahkan pada kesalahan besar. Hal

ini disebabkan oleh istilah kuadrat, secara eksponensial meningkatkan kesalahan yang lebih besar dibandingkan dengan yang lebih kecil[19].

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

Root Mean Squared Error adalah akar kuadrat dari kesalahan kuadrat rata-rata. Ini adalah salah satu metrik evaluasi yang paling populer karena Root Mean Squared Error dapat ditafsirkan dalam unit yang sama dengan vektor respons atau unit Y, sehingga mudah untuk menghubungkan informasinya[19].

$$\text{RAE} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{\sum_{i=1}^n |y_i - \bar{y}|}$$

Relative absolute error, juga dikenal sebagai jumlah sisa kuadrat, di mana bilah Y adalah nilai rata-rata Y, mengambil kesalahan absolut total dan menormalkannya. Dengan membaginya dengan kesalahan absolut total dari prediktor sederhana.

$$\text{RSE} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Relative squared error sangat mirip dengan relative absolute error, tetapi diadopsi secara luas oleh komunitas ilmu data karena digunakan untuk menghitung R-kuadrat.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

R-kuadrat bukan kesalahan per kata tetapi merupakan metrik populer untuk keakuratan model. Ini mewakili seberapa dekat nilai data dengan garis regresi yang dipasang. Semakin tinggi R-kuadrat, semakin baik modelnya sesuai dengan data.

Interpretasi R^2 :

- $R^2 = 1$ berarti model prediksi sempurna.
- $R^2 = 0$ berarti model tidak lebih baik dari rata-rata.

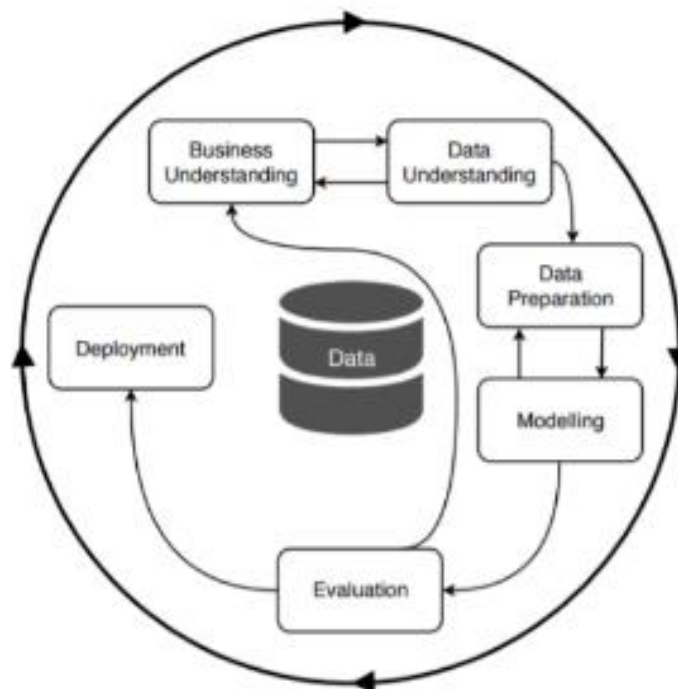
- Nilai negatif R^2 berarti model lebih buruk daripada hanya menggunakan rata-rata sebagai prediksi.

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%$$

MAPE mengukur kesalahan absolut rata-rata sebagai persentase dari nilai aktual. Ini sering digunakan karena memberikan kesalahan dalam persentase yang mudah dipahami.

2.12 Cross-Industry Standard Process for Data Mining (*CRISP-DM*)

CRISP-DM (Cross-Industry Standard Process for Data Mining) adalah sebuah pendekatan atau metodologi yang digunakan untuk mengelola dan menjalankan proyek data mining secara sistematis. Metodologi ini menyediakan panduan langkah demi langkah untuk melaksanakan proyek data mining, mulai dari pemahaman bisnis hingga evaluasi hasil[23].



Gambar 9 Arsitektur CRISP-DM

CRISP-DM (Cross-Industry Standard Process for Data Mining) adalah metodologi yang digunakan secara luas dalam industri untuk mengelola proyek-proyek analisis data, khususnya dalam konteks data mining dan data science. Dikembangkan pada akhir 1990-an, CRISP-DM memberikan panduan yang terstruktur bagi para profesional dalam setiap tahap dari proses analisis data, mulai dari pemahaman masalah hingga implementasi solusi. Metodologi ini menjadi standar industri yang umum digunakan, memberikan kerangka kerja yang kokoh untuk mengatasi kompleksitas yang muncul dalam proyek-proyek analisis data di berbagai bidang industri. CRISP-DM terdiri dari enam tahap utama, yaitu **Pemahaman Bisnis (Business Understanding)**, yang merupakan tahap awal proyek data mining dan berfokus pada pemahaman tujuan bisnis serta menentukan kebutuhan informasi yang harus dipenuhi melalui analisis data. Selanjutnya, tahap **Pemahaman Data (Data Understanding)** bertujuan untuk memahami data secara mendalam melalui eksplorasi awal guna mengetahui karakteristik, kualitas, dan potensi informasi yang terkandung. Tahap berikutnya adalah **Persiapan Data (Data Preparation)**, di mana data mentah yang dikumpulkan diolah dan disesuaikan agar siap digunakan dalam analisis dan pemodelan. Pada tahap **Pembuatan Model (Modeling)**, berbagai teknik dan algoritma digunakan untuk membangun model prediktif berdasarkan data yang telah dipersiapkan sebelumnya guna mendukung pengambilan keputusan atau membuat prediksi. Tahap **Evaluasi (Evaluation)** bertujuan untuk menilai kualitas dan performa model yang dibangun dengan menggunakan metrik-metrik evaluasi relevan sehingga memastikan model dapat memberikan hasil yang akurat dan berguna dalam memecahkan masalah bisnis. Terakhir, tahap **Penerapan (Deployment)** melibatkan implementasi model yang telah dievaluasi ke lingkungan produksi atau operasional sehingga dapat digunakan secara nyata untuk menyelesaikan masalah bisnis yang dihadapi.

2.13 Flask

Flask adalah sebuah microframework web yang ditulis dalam bahasa pemrograman Python yang dirancang untuk mempermudah pengembangan aplikasi web melalui penyediaan alat dan pustaka esensial tanpa memaksakan arsitektur atau pola tertentu. Flask memungkinkan pengembang untuk membangun aplikasi web dengan struktur yang sederhana namun mudah diperluas melalui berbagai ekstensi. Popularitasnya di kalangan pengembang juga didukung oleh komunitas yang aktif dan dokumentasi yang komprehensif, menjadikan Flask pilihan ideal untuk proyek-proyek skala kecil hingga menengah yang memerlukan pengembangan cepat dan prototipe, tanpa memerlukan kompleksitas bawaan dari framework yang lebih besar seperti Django [31].

2.14 Penelitian Terkait

Penelitian ini menggunakan beberapa studi literatur sebagai referensi untuk menambah pemahaman dan mengembangkan wawasan baru dari penelitian sebelumnya. Beberapa penelitian penting yang menjadi dasar pembuatan penelitian ini terdapat pada tabel 2.

Tabel 2 Penelitian Terkait

No.	Judul	Peneliti / Tahun	Fitur	Target	Metode	Hasil
1	<i>A Hybrid Air Quality Index Prediction Model Based on CNN and Attention Gate Unit[7]</i>	Jingyang Wang, dkk. (2022)	Data kualitas udara dan meteorologi per jam dari 1 Januari 2018 hingga 30 Juni 2021 di Shijiazhuang, Hebei, China, diperoleh dari situs Shanghai Qingyue Data dan NowAPI, dengan total 39.408 data. Berdasarkan standar China, polutan utama dalam AQI meliputi SO ₂ , O ₃ , PM _{2.5} , NO ₂ , PM ₁₀ , dan CO. Faktor meteorologi seperti curah hujan, suhu, kelembaban, angin, dan cuaca memengaruhi AQI dengan	AQI	<i>CNN dan AGU(Attention Gate Units)</i>	<ul style="list-style-type: none"> - Model prediksi AQI berbasis CNN-AGU yang diusulkan dalam penelitian ini lebih unggul dibandingkan sembilan model lainnya(<i>MLR, RFR, SVR, LSTM, GRU, BD-LSTM, AGU, CNN-LSTM, AGU, CNN-GRU</i>). - CNN berperan mengekstrak fitur penting dari data.

			mendukung penyebaran dan pengenceran polutan, sehingga ditambahkan dalam input prediksi AQI.			<ul style="list-style-type: none"> - Attention Mechanism yang tertanam dalam AGU meningkatkan kemampuan memori unit gated untuk memproses data. - DAM (Data Adjustment Module) pada AGU meningkatkan sensitivitas unit gated terhadap pembelajaran data historis.
2	Prediksi Indeks Standar Pencemar Udara Menggunakan Metode Long Short-Term Memory	Benny Khalid Hidayatullah, dkk. (2022)	Rentang waktu data mencakup 1 Januari 2015 hingga 31 Oktober 2021, dengan total 12.480 entri dengan variabel tanggal, stasiun, PM_{10} , SO_2 , NO_2 , CO , O_3 .	PM_{10} , SO_2 , NO_2 , CO , O_3	LSTM	Hasil pengujian parameter terbaik dari kelima zat tersebut menghasilkan rentang nilai $RMSE$ dari 0.0431 sampai 0.07237 dan skor R^2 dari rentang nilai 0.58157 hingga 0.88621.

	Berbasis Web (Studi Kasus Pada Kota Jakarta)[8]					
3	<i>Prediction of Water Levels on Peatland using Deep Learning.</i> [12]	Namora, Jan Everhard Riwurohi. (2022).	Data tinggi muka air gambut dari Desember 2018 hingga November 2021 diperoleh melalui telemetri water logger milik BRGM yang dikelola oleh BPPT. Data mentah per jam ini diolah menjadi data harian melalui agregasi rata-rata.	<i>Water Level</i>	Dalam penelitian ini, diusulkan untuk menggunakan metode pembelajaran mendalam Long Short-Term Memory (LSTM), dan menerapkan metodologi <i>CRISP-DM</i> .	Dari kesimpulan penelitian, dapat disimpulkan bahwa model LSTM mampu memprediksi tingkat air dengan baik, seperti yang ditunjukkan oleh rata-rata <i>RMSE</i> sebesar 0,07 m, rata-rata R^2 sebesar 0,85, dan rata-rata <i>MAE</i> sebesar 0,04 m. Parameter-model LSTM yang optimal mencakup 50 epoch, rasio data latih terhadap data uji sebesar 70%:30%, dan 2 lapisan tersembunyi.
4	Pengujian Algoritma <i>Long short term memory</i> untuk	Ali Khumaidi, dkk. (2022).	Dataset yang digunakan mencakup 9.231 catatan, mulai dari tanggal 19 Mei 2019 pukul 00:04:19 hingga 18 Agustus	PM_{10} , ISPU, suhu, dan kelembaban.	Pemodelan LSTM dilakukan dengan 4 <i>hidden layer</i> , menggunakan batch size sebesar 32, optimizer	Prediksi suhu memiliki <i>RMSE</i> dengan Train Score 2,37 dan test score 3,15, menandakan kinerja yang cukup baik.

	prediksi kualitas udara dan suhu kota bandung[24]		2019 pukul 23:52:37. Terdapat 21 parameter dalam data ini, yaitu profil, tanggal, jam, O ₃ , CO, NO ₂ , SO ₂ , NO, CO ₂ , VOC, PM1, PM2.5, PM4, PM10, TSP, suhu (TEMP), kelembapan (HUM), kecepatan angin (WS), arah angin (WD), ISPU, dan status.		adalah Adam, dilatih selama 1000 epoch, dan fungsi <i>loss</i> yang digunakan adalah mean squared error.	Prediksi PM_{10} menunjukkan <i>RMSE</i> dengan Train Score 1,92 dan test score 1,76, namun kinerjanya kurang baik karena test score lebih tinggi dari Std test (1,75). Prediksi kelembapan memiliki <i>RMSE</i> dengan Train Score 7,56 dan test score 6,97, menandakan kinerja yang cukup baik. Sedangkan prediksi ISPU memiliki <i>RMSE</i> dengan Train Score 1,89 dan test score 1,85, juga menunjukkan kinerja yang cukup baik.
5	Implementasi Metode <i>Long short term memory</i> (LSTM) untuk Memprediksi Harga Bahan	Jumina Cahyani, dkk. (2023).	32 atribut yang merupakan harga bahan pokok nasional dengan total data sebanyak 903 data.	Prediksi pada 4 harga bahan pokok dilakukan untuk 5 hari kedepan. Yaitu, harga daging ayam ras segar,	Metode <i>Long short term memory</i> (LSTM) sering digunakan karena kemampuannya dalam menjaga informasi jangka panjang. Oleh karena itu, LSTM digunakan untuk	Dalam pengoptimalan hasil prediksi, tiga metode optimasi digunakan sebagai skenario uji coba, yaitu Adaptive Gradient (ADAGRAD), Adaptive Moment Gradient (ADAM), dan Root Mean

	Pokok Nasional[11]			harga beras kualitas bawah II, harga minyak goreng, dan harga minyak goreng curah.	memprediksi harga bahan pokok nasional.	Square Error (RMSPProp). Berdasarkan uji coba tersebut, model terbaik untuk harga beras kualitas bawah II ditemukan menggunakan metode ADAM, dengan nilai evaluasi <i>RMSE</i> sebesar 0.0492 dan <i>R</i> ² Score sebesar 0.8852. Untuk harga daging ayam, model terbaik juga menggunakan metode ADAM, dengan nilai evaluasi <i>RMSE</i> sebesar 0.0937 dan <i>R</i> ² Score sebesar 0.5949. Sedangkan untuk harga minyak goreng, model terbaik ditemukan menggunakan metode RMSProp, dengan nilai evaluasi <i>RMSE</i> sebesar 0.0313 dan <i>R</i> ² Score sebesar 0.7492.
--	--------------------	--	--	--	---	---

6	<i>A deep learning approach for prediction of air quality index in a metropolitan city</i> [25]	R. Janarthanan, dkk (2021).	Particulate Matter (PM2.5 dan PM10), Ozon (O ₃), Sulfur Dioksida (SO ₂), Nitrogen Dioksida (NO ₂), Karbon Monoksida (CO), Timbal (Pb), dan Amonia (NH ₃).	AQI	Model LSTM (Long Short-Term Memory): Digunakan untuk memprediksi nilai AQI (Indeks Kualitas Udara) dengan perbandingan berbagai polutan seperti PM _{2.5} , NO ₂ , SO ₂ , CO, dan Ozon.	Model Proposed (LSTM) menunjukkan performa terbaik dengan RMSE terendah (10.9) dan korelasi tertinggi (R = 0.97), mengungguli semua model lain, termasuk EMD-CNN (RMSE = 46.26, R = 0.81).
7	<i>Research on a hybrid model for cooling load prediction based on wavelet threshold denoising and deep learning: A study in China.</i> [26]	Fuyu Wang, dkk. (2022)	Data historis beban pendinginan, suhu, kelembapan, dan jumlah chiller dipilih untuk membentuk kumpulan data sampel.	<i>cooling load (beban pendinginan)</i>	Model prediksi beban pendingin WTD-CNN-LSTM menggunakan denoising WTD untuk membersihkan data, CNN untuk mengekstraksi fitur kompleks dari variabel input, dan LSTM untuk memahami hubungan waktu jangka panjang, sehingga menghasilkan prediksi yang lebih akurat. Selain itu, dalam penelitian ini, Untuk	Model prediksi beban pendingin WTD-CNN-LSTM yang diusulkan dapat secara efektif menghilangkan noise dan mempertahankan fitur data untuk meningkatkan akurasi prediksi, serta mengungguli model lain dengan pengurangan MAE sebesar 8.51–206.98 kW, RMSE sebesar 12.33–162.01 kW, dan MAPE sebesar 0.23%–11.08% pada berbagai

					memverifikasi keunggulan model yang diusulkan, dilakukan eksperimen komparatif yang berbeda dengan SVR, CNN, RNN, LSTM, GRU, BiLSTM dan CNN-LSTM	interval waktu (10 menit, 30 menit, 1 jam), menunjukkan kinerja terbaik dan kemampuan generalisasi yang lebih baik.
8	Perbandingan Model RNN, Model LSTM, dan Model GRU dalam Memprediksi Harga Saham-Saham LQ45[27]	Andrew Nilsen (2022)	Data sekunder yang diperoleh dari website Yahoo Finance yakni, data harga saham-saham yang terdaftar di indeks LQ45 dengan rentang waktu dari awal saham tersebut terdaftar di Bursa Saham Indonesia hingga waktu dibuatnya penelitian ini yaitu 20 April 2022.	Masing- Masing Saham LQ45	Metode yang digunakan adalah perbandingan model <i>Recurrent Neural Network</i> (RNN), Long Short-Term Memory (LSTM), dan Gated <i>Recurrent Unit</i> (GRU) dalam memprediksi harga saham.	Hasil penelitian menunjukkan bahwa model GRU secara konsisten memiliki nilai galat yang lebih rendah (<i>RMSE</i> , <i>MSE</i> , dan <i>MAE</i>) dibandingkan dengan model RNN dan LSTM untuk sebagian besar saham yang terdaftar di indeks LQ45.
9	Perbandingan Algoritma Linear Regression, LSTM, Dan GRU Dalam Memprediksi	Khalis Sofi, dkk (2021)	Variabel data berisi harga previous saham, harga open, harga high, harga low, harga close, dan volume. data publik yang diambil dari https://www.idx.co.id/ dengan	Harga saham KEJU	Penelitian ini bertujuan untuk memprediksi harga saham dengan membandingkan algoritma Linear Regression, Long Short-Term Memory	Setelah melalui proses training dan testing, hasil analisis menunjukkan bahwa algoritma Gated <i>Recurrent Unit</i> (GRU) memiliki performa terbaik dalam

	Harga Saham Dengan Model Time Series[28]		data penjualan saham dari tanggal 25 November 2019 sampai dengan 08 Juni 2021.		(LSTM), dan Gated <i>Recurrent Unit</i> (GRU) menggunakan dataset publik dari Indonesia Stock Exchange (IDX).	memprediksi harga saham dibandingkan dengan Linear Regression dan Long Short-Term Memory (LSTM). Hal ini dibuktikan dengan nilai galat GRU yang paling rendah, yaitu <i>RMSE</i> sebesar 0.034, <i>MSE</i> sebesar 0.001, dan <i>MAE</i> sebesar 0.024.
10	Prediksi Kualitas Udara Dengan Metoda LSTM, <i>Bidirectional LSTM</i> , dan GRU[29]	Yadi Karyadi, Handri Santoso (2022)	Suhu, PM_{10} , Kelembapan dan ISPU tahun 2019.	suhu, kelembaban, PM_{10} , dan ISPU.	Long Short-Term Memory (LSTM), <i>Bidirectional Long Short-Term Memory (LSTM-Bi)</i> , dan Gated <i>Recurrent Unit (GRU)</i> dengan pembagian data 75% training dan 25% testing.	Hasil penelitian menunjukkan bahwa model LSTM dan LSTM-Bi memiliki performa yang lebih baik dibandingkan GRU, khususnya pada prediksi variabel suhu, kelembaban, dan ISPU. Pada model LSTM, nilai <i>RMSE</i> untuk suhu adalah 3.18 (dibandingkan dengan standar deviasi 4.40), kelembaban 6.96 (dibandingkan dengan 19.32), dan ISPU 1.84

						(dibandingkan dengan 1.90). Sedangkan pada model LSTM-Bi, nilai RMSE untuk suhu adalah 3.23, kelembaban 7.07, dan ISPU 1.86. Model GRU tidak memberikan hasil yang sebaik LSTM dan LSTM-Bi untuk data time series kualitas udara.
11	<i>Daily scale air quality index forecasting using Bidirectional recurrent neural networks: Case study of Delhi, India</i> [30]	Chaitanya Baliram Pande, dkk. (2024)	PM _{2.5} , PM ₁₀ , NO ₂ , NO _x , SO ₂ , CO, dan O ₃ . Kumpulan data mengungkapkan 3318 pengamatan yang mencakup periode dari 1 Januari 2014 hingga 31 Desember 2022.	AQI	Penelitian ini menggunakan berbagai model <i>deep learning</i> termasuk Long-Short Term Memory (LSTM), <i>Bidirectional LSTM (Bi-LSTM)</i> , <i>Bidirectional Recurrent Neural Networks (Bi-RNN)</i> , serta model regresi Kernel Ridge Regression (KRR) untuk memprediksi indeks kualitas udara harian (AQI) di kota Delhi, India.	Hasil penelitian menunjukkan bahwa model Bi-RNN secara konsisten mengungguli model lainnya baik pada fase pelatihan maupun pengujian, dengan mencapai koefisien determinasi maksimum ($R^2 = 0.954$) dan root mean square error minimum ($RMSE = 25.755$), sementara model KRR menunjukkan kinerja terlemah.

Beberapa penelitian sebelumnya telah menunjukkan keunggulan berbagai algoritma dalam memprediksi kualitas udara menggunakan *deep learning*. Dalam konteks prediksi time series, seperti yang dilakukan oleh Ali Khumaidi dkk. (2022), algoritma LSTM menunjukkan performa yang baik dalam memprediksi suhu dan kelembaban, dengan hasil yang akurat dan konsisten. Inovasi dalam model prediksi kualitas udara juga menjadi fokus utama penelitian ini. Jingyang Wang dkk. (2022) menekankan pentingnya model hybrid seperti CNN-AGU yang mampu mengekstraksi fitur penting dari data dan meningkatkan kemampuan memori unit gated. Mengikuti jejak ini, dalam penelitian ini mengimplementasikan teknik *Bidirectional* pada LSTM dan GRU serta RNN untuk memproses informasi dari kedua arah (masa lalu dan masa depan), serta menggunakan teknik *Dropout* untuk mengurangi overfitting. Pendekatan ini diharapkan dapat meningkatkan akurasi prediksi kualitas udara.

Hasil penelitian menunjukkan bahwa LSTM dan LSTM-*Bidirectional* secara konsisten memberikan hasil prediksi yang lebih baik dibandingkan GRU, terutama dalam memprediksi variabel suhu, kelembaban, dan Indeks Standar Pencemar Udara (ISPU). Temuan ini sejalan dengan penelitian Benny Khalid Hidayatullah dkk. (2022) dan Yadi Karyadi serta Handri Santoso (2022), yang juga menemukan bahwa LSTM unggul dalam prediksi berbagai polutan. Konsistensi hasil ini memperkuat validitas penggunaan LSTM dan teknik *Bidirectional* dalam prediksi kualitas udara.

Selain itu, penelitian ini juga berfokus pada aplikasi praktis dari model prediksi yang dikembangkan. Dengan menggunakan Flask, hasil prediksi dapat diakses melalui aplikasi berbasis web, memungkinkan masyarakat untuk memantau kualitas udara secara real-time. Pendekatan ini sejalan dengan penelitian Chunhao Liu dkk. (2023) yang menekankan pentingnya dukungan model prediksi terhadap kelompok rentan dan peringatan dini terhadap peristiwa kualitas udara yang buruk. Dengan demikian, penelitian ini tidak hanya meningkatkan akurasi prediksi tetapi juga menyediakan alat praktis untuk pemantauan kualitas udara, mendukung upaya pengendalian polusi di kota-kota besar.

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat

Penelitian dan pembuatan skripsi ini adalah dilakukan pada :

Waktu : Januari 2024 – Desember 2024.

Tempat : Laboratorium Teknik Komputer Universitas Lampung.

Jadwal kegiatan penelitian yang dilakukan adalah sebagai berikut:

Tabel 3 Jadwal Kegiatan Penelitian

No.	Aktifitas	BULAN											
		1	2	3	4	5	6	7	8	9	10	11	12
1	Pemahaman Bisnis	■	■	■									
2	Pemahaman Data dan Persiapan Data			■	■	■							
3	Pembuatan Model dan Evaluasi				■	■	■	■					
4	<i>Deployment</i>						■	■					
5	Laporan Skripsi	■	■	■	■	■	■	■	■	■	■	■	■

3.2 Alat dan Bahan

Alat dan bahan yang dibutuhkan dalam penelitian ini adalah sebagai berikut:

3.2.1 Alat

- Perangkat keras

Laptop *Acer Aspire 5*, VGA AMD Ryzen 3 5300U with *Radeon Grapich*, RAM 8GB DDR4, SSD 512GB, SDD 128GB digunakan sebagai media perancangan dan pengembangan algoritma forecasting yang diteliti serta sebagai perangkat pembuatan laporan penelitian.

- Perangkat lunak (*Software*)

Berikut merupakan perangkat lunak yang digunakan dalam penelitian ini beserta penjelasan terkait kegunaannya.

Tabel 4 Software Pengembangan Deep learning.

No	Nama <i>Software</i>	<i>Version</i>	Kegunaan
1	<i>Python</i>	3.11.5 (conda)	Bahasa pemrograman yang digunakan untuk pengembangan perangkat lunak dan analisis data.
	<i>a. Numpy</i>	1.24.3	Library Python untuk komputasi numerik, termasuk operasi array dan matriks.
	<i>b. Pandas</i>	2.0.3	Library Python untuk manipulasi dan analisis data, terutama dalam bentuk tabel.
	<i>c. Matplotlib</i>	3.7.2	Library Python untuk visualisasi data dalam bentuk grafik dan plot.
	<i>d. Seaborn</i>	0.12.2	Library Python untuk visualisasi data statistik yang dibangun di atas Matplotlib.
	<i>e. Scikit-learn</i>	1.3.0	Library Python untuk <i>machine learning</i> dan data mining yang menyediakan algoritma klasifikasi, regresi, klustering, dan lainnya.
	<i>f. Tensorflow</i>	2.15.0	Library Python yang digunakan untuk pembelajaran mesin dan <i>deep learning</i> .
	<i>g. Keras</i>	2.15.0	High-level <i>neural networks</i> API yang berjalan di atas TensorFlow, Theano, atau Microsoft Cognitive Toolkit.

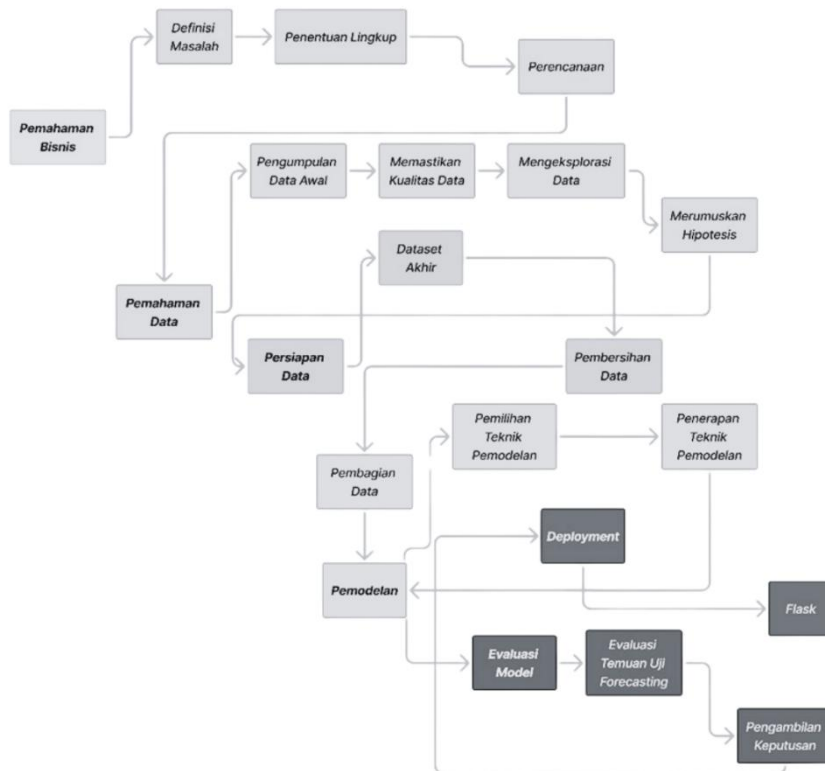
	<i>h. H5 (HDF5)</i>	3.9.0	Format file untuk penyimpanan dan pertukaran data ilmiah, khususnya data besar dan kompleks.
	<i>i. requests</i>	2.31.0	Modul yang digunakan untuk membuat permintaan HTTP dalam Python.
	<i>j. os</i>	3.11.5	Modul yang menyediakan berbagai fungsi untuk berinteraksi dengan sistem operasi.
2	<i>Postman</i>	11.0.7	Aplikasi untuk mengirim permintaan dan menerima balasan dari API web.
3	<i>Visual Studio Code</i>	1.87.2	Code editor yang dikembangkan oleh Microsoft, yang mendukung berbagai bahasa pemrograman.
	<i>Jupyter notebook</i>	6.5.4	Aplikasi web open-source yang memungkinkan pembuatan dan berbagi dokumen yang berisi kode live, persamaan, visualisasi, dan narasi teks.
4	<i>Flask</i>	3.0.3	Sebuah framework web mikro untuk bahasa pemrograman Python.

3.2.2 Bahan

Data yang digunakan pada penelitian ini adalah data kualitas udara di salah satu perwakilan wilayah di Provinsi Lampung untuk melihat tren di 15 kota/ kabupaten di Provinsi Lampung. Data berasal dari situs weatherbit.io dalam kurun waktu 13 januari 2024 dengan interval pengukuran setiap jam. Dataset terdiri dari 11 kolom yang mencakup berbagai variabel seperti AQI, CO, datetime, NO_2 , O_3 , PM_{10} , $PM_{2.5}$, SO_2 , timestamp_local, timestamp_utc, dan ts. Beserta tambahan data cuaca berupa data kecepatan angin, temperature, relative humidity dan pressure juga berasal dari API weather bit terkait cuaca. Setiap kolom memiliki total 17.976 baris data.

3.3 Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini menggunakan langkah langkah yang ada dalam metode *CRISP-DM*. Tahapan tersebut terdiri dari 6 tahapan. Di dalam tahapan tersebut terdapat beberapa sub unit yang berisi inti dalam pembahasan ini. Sub unit tersebut ditampilkan dalam gambar [14](#).



Gambar 10 Flowchart Penelitian

CRISP-DM (Cross-Industry Standard Process for Data Mining) terdiri dari enam tahap yang dilakukan secara berulang. Tahap pertama adalah pemahaman bisnis, yang mencakup definisi masalah, penentuan lingkup, dan perencanaan. Tahap kedua adalah pemahaman data, yang meliputi pengumpulan data awal, memastikan kualitas data, mengeksplorasi data, dan merumuskan hipotesis. Selanjutnya, tahap ketiga melibatkan persiapan data, di mana dataset akhir disusun dari data mentah. Tahap keempat adalah tahap pemodelan, di mana teknik pemodelan yang sesuai dipilih dan diterapkan. Pada Tahap kelima, evaluasi model dilakukan, di mana temuan dievaluasi dan keputusan diambil berdasarkan hasil tersebut. Akhirnya,

dalam tahap implementasi, model-model tersebut diimplementasikan untuk penggunaan praktis[18]. Berikut adalah langkah-langkah penelitian menggunakan metode *CRISP-DM* untuk memprediksi kualitas udara menggunakan :

1. Pemahaman Bisnis (*Business Understanding*)

Pertama, definisikan masalah yang dipecahkan dalam penelitian ini. Masalah yang diselesaikan sesuai dengan rumusan masalah, antara lain.

1. Kota/kabupaten mana di Provinsi Lampung yang paling representatif sebagai sumber dataset untuk melatih model deep learning melalui eksplorasi data analisis?
2. Bagaimana tingkat akurasi dari model deep learning yang digunakan dalam pelatihan?
3. Bagaimana tingkat akurasi dari model deep learning yang diterapkan menggunakan data langsung dari ke-15 kota di Provinsi Lampung?

Kemudian, penentuan lingkup yang merupakan batasan dari permasalahan yang diambil, antara lain.

1. Target prediksi adalah pemantauan kualitas udara (*AQI*).
2. Data yang digunakan untuk pelatihan model merupakan data history pada salah satu kota di Provinsi Lampung. Data diperoleh menggunakan *API Weatherbit*. Menimbang keterbatasan dalam pengambilan data, maka tren yang dapat ditampilkan hanya 2 tahun terakhir.
3. Arsitektur yang digunakan merupakan arsitektur *classical*, *Double*, *Double* dengan penambahan regulasi *Dropout* dan *Bidirectional*. Hyperparameter yang digunakan adalah menggunakan timesteps 24 dikarenakan kebutuhan data PM_{25} dan PM_{10} selama 24 jam untuk menghitung *AQI*, *epochs* 50 dan 100, *learning rate* 0.001 dan 0.005, *neurons* 50 dan 100 dan regulasi *Dropout* 0.2.
4. Evaluasi kinerja model dilakukan dengan memeriksa indikator berikut:, *Mean Squared Error (MSE)* dan *Koefisien Determinasi (R^2)*.

5. Pengembangan model dan visualisasi hasil dilakukan menggunakan framework flask secara lokal menggunakan localhost untuk menampilkan prediksi wilayah lainnya.

Kemudian pada fase perancangan, sesuai dengan tujuan penelitian yang telah ada.

1. Menggunakan salah satu kota untuk dijadikan dataset dalam pembuatan model deep learning.
2. Mendapatkan akurasi dan kelebihan serta kekurangan dari model deep learning yang digunakan.
3. Mendapatkan akurasi dari model deep learning yang digunakan.

Pada tahap ini, didapatkan hasil perencanaan yang berkaitan dengan tujuan penelitian dan menjawab rumusan dari masalah yang ada.

2. Pemahaman Data (*Data Understanding*)

Pada tahap pemahaman data terdapat beberapa inti pembahasan didalamnya. Langkah pertama yang perlu dilakukan adalah pengumpulan data yang berkaitan dengan kualitas udara. Data yang dikumpulkan adalah data dari ke-15 kabupaten/kota di provinsi lampung. Data diambil dari salah satu wilayah yang mewakili kabupaten/kota tersebut. Pada tahap ini dilakukan eksplorasi data untuk melihat tren kualitas udara di provinsi lampung. Berdasarkan tren tersebut, diambil salah satu wilayah yang relevan sebagai dataset untuk melakukan pembuatan model prediksi.

Kemudian pada tahap selanjutnya adalah memastikan kualitas data dengan menggunakan korelasi pearson. Setelah itu, eksplorasi data yang telah dipastikan kualitasnya dengan melakukan eksplorasi data analisis (EDA). Kemudian buat sebuah hipotesis berdasarkan penelitian penelitian sebelumnya yang terkait dengan penelitian ini dan data yang telah dieksplorasi. Tujuan dari tahap ini adalah untuk mengeksplorasi data yang dapat digunakan dan merumuskan hipotesis berdasarkan penelitian penelitian terkait.

Pada tahap ini, dilakukan analisa tren terhadap data kualitas udara di seluruh wilayah Provinsi Lampung. Kemudian berdasarkan hipotesis terkait penelitian terkait, diambil satu data wilayah yang digunakan untuk data pelatihan. Data

wilayah tersebut ditambahkan sehingga memiliki data polutan ($PM_{2.5}$, PM_{10} , SO_2 , NO_2 , CO , dan $Ozone$) dan data meteorologi (wind speed, temperature, relative humidity dan pressure). Kemudian data tersebut dipastikan menggunakan korelasi pearson untuk mendapatkan fitur data yang dapat digunakan sebagai data pelatihan untuk pembuatan model. Berdasarkan korelasi data tersebut, diperoleh hipotesis bahwa data tersebut dapat memprediksi kualitas udara di masa depan.

3. Persiapan Data (*Data Preparation*)

Pada tahap persiapan data, sudah didapatkan kesimpulan berdasarkan analisa tren untuk menggunakan data salah satu wilayah untuk dijadikan data pelatihan model sehingga didapatkan dataset akhir setelah melalui pemahaman data. Data yang ada dibersihkan apabila terdapat data kosong ataupun data terduplikat. Hasil dari tahap ini adalah berupa dataset yang siap digunakan untuk melakukan pemodelan.

Pada tahap ini ditetapkan terkait data fitur dan data target. Data fitur yang digunakan bergantung pada hubungan korelasi pearson data fitur dengan data AQI yang merupakan data target. Data fitur yang digunakan diharuskan memiliki hubungan sangat kuat dengan data target untuk meminimilasi kesalahan. Kemudian, data fitur masuk pada fase *preprocessing* untuk dilakukan pengecekan data kosong dan data duplikasi agar bersih dengan melakukan pembersihan data. Setelah itu, data di normalisasi untuk memudahkan pengolahan data menggunakan MinMaxScaler dengan melakukan skala data antara 0 dan 1. Hal yang sama dilakukan pada target, tetapi karena data target memiliki data satu dimensi, diperlukan perubahan data menjadi 2 dimensi untuk dapat menggunakan MinMaxScaler. Proses perubahan menjadi 2 dimensi menggunakan metode reshape agar menjadi 2 dimensi. Setelah data di normalisasi, data diubah menjadi data sekuensial dengan bentuk tiga dimensi. Dimana dalam bentuk tersebut berisi data fitur yang telah diskala, target dan jumlah waktu data fitur yang diambil (timestep). Dalam hal ini, timestep yang digunakan untuk mengambil data fitur adalah 24 jam. Hal ini dikarenakan untuk mendapatkan AQI, dibutuhkan nilai rata rata polutan yang digunakan untuk perhitungan AQI. Setelah itu, untuk membuat pelatihan model, diperlukan pembagian data antara data latih dan data uji untuk menguji akurasi selama pelatihan. Dalam penelitian ini, digunakan train test split untuk membagi data latih

dan data uji dengan perbandingan 80% sebagai data latih dan 20% sebagai data uji beserta penggunaan random state sebanyak 32. Setelah melalui proses preprocessing data, data latih dan data uji telah siap untuk melakukan pemodelan.

4. Pembuatan Model (*Modeling*)

Pada tahap ini, terdapat 2 (dua) inti pembahasan, yakni tahap pemilihan teknik dan penerapan teknik pemodelan. Pada tahap pemilihan teknik, teknik yang digunakan adalah metode *deep learning* yang dapat mengolah data sekuensial, antara lain adalah RNN, LSTM dan GRU.

Pada tahap ini, dilakukan uji terhadap beberapa arsitektur RNN, LSTM dan GRU. Arsitektur yang digunakan antara lain adalah *Single*, *Double*, *Double* dengan *Dropout* dan *Double* dengan teknik *Bidirectional*. Hyper parameter yang digunakan dibatasi dengan jumlah epoch 50 dan 100, jumlah *neuron* 50 dan 100, *learning rate* 0.001 dan 0.0005 serta jumlah timestemp sequence adalah 24.

5. Evaluasi (*Evaluation*)

Untuk menguji akurasi model yang telah dilatih, pengecekan error dilakukan pada tahap evaluasi. Model dievaluasi menggunakan *MSE* dan R^2 untuk melihat akurasi dalam evaluasi model. Hasil dari tahap ini merupakan metrik akurasi dari masing masing teknik model yang telah dilatih. Metrik akurasi dievaluasi kembali untuk dilakukan pengambilan keputusan agar dapat diterapkan dalam tahap deployment. Setelah melalui proses evaluasi, model dipilih 3 model terbaik yang mewakili masing masing model.

6. Penerapan (*Deployment*)

Fase Penerapan berkaitan dengan menerapkan hasil proyek penambangan data untuk memastikan bahwa mereka tersedia dan melayani kebutuhan bisnis pengguna. Setelah model disimpan, maka model diterapkan pada web menggunakan *flask*. Model yang telah dievaluasi dan disimpan dipanggil menggunakan library tensorflow dengan menggunakan kode *load_model*. Pengujian visualisasi menggunakan flask dilakukan secara lokal menggunakan localhost. Hasil akhir dari tahap ini adalah tampilan hasil prediksi menggunakan model yang telah dibuat pada beberapa lokasi di wilayah lampung.

V. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Pada penelitian ini dibahas mengenai bagaimana implementasi Algoritma *deep learning* untuk memprediksi AQI di provinsi lampung menggunakan model RNN, LSTM, dan GRU dengan implementasi data langsung untuk memberikan informasi tentang bagaimana kualitas udara di Provinsi Lampung. Berikut merupakan beberapa kesimpulan berdasarkan penelitian yang telah dilakukan.

1. Berdasarkan tren data selama 2 tahun di ke-15 kota/kabupaten di provinsi lampung dengan memperlihatkan tren aqi selama Covid 19 selama 18 bulan sebelum ditetapkan berakhir pada 21 Juni 2023 hingga akhir desember 2023. Rata rata AQI di Provinsi Lampung berada di kisaran **40 – 77 (Good (0-50) – Moderate (51 -100))**, dengan rata rata AQI tertinggi adalah kabupaten Pringsewu dan Lampung Utara dengan nilai 77 (dibulatkan). Sedangkan standar deviasi data Pringsewu lebih besar dari pada Lampung Utara ($33 > 32$) sehingga digunakan data Pringsewu sebagai sumber dataset untuk pelatihan model deep learning.
2. Berdasarkan korelasi pearson, variabel PM_{25} dan PM_{10} memiliki korelasi yang sangat kuat dengan AQI. Dengan ke-2 input variabel tersebut pelatihan model prediksi memiliki rata rata R^2 diatas 90% dengan model terbaik adalah Model **LSTM Double dengan 100 neurons, 100 epochs, dan 0.0005 learning rate** yang memiliki nilai R^2 tertinggi yaitu 0.9360 dan MSE 0.0011. Sedangkan RNN (0.9246, 0.0012) dan GRU (0.9325, 0.0011). Meskipun memiliki nilai akurasi tertinggi tetapi model LSTM memiliki waktu pemrosesan testing selama 9ms/step sebanyak 112 step dengan total params 121,701 (475.39 KB). Lebih lambat dari model RNN selama 5ms/step dengan total params 30,501 (119.14 KB) dan Lebih cepat dari model GRU selama 19ms/step dengan total params 91,901 (358.99 KB) dengan pengaturan hyperparameter tuning yang sama.
3. Tiga arsitektur model terbaik dari masing masing model adalah RNN (0.9314, 0.0011), LSTM (0.9360, 0.0011), dan GRU (0.9355, 0.0011) dengan konfigurasi masing-masing: RNN Double (50 neurons, 100 epochs, lr 0.001), LSTM Double (100 neurons, 100 epochs, lr 0.0005), dan Bi-GRU (100

neurons, 100 epochs, lr 0.001). Ketiga model tersebut yang digunakan dalam tahapan deployment.

4. Dengan dataset berjumlah 17.902 (2 tahun), AQI dapat diprediksi menggunakan parameter $PM^{2.5}$ dan PM^{10} sebanyak 24 nilai (satu hari). Tingkat akurasi R^2 pada **kabupaten Pringsewu**, wilayah sub-urban, untuk model RNN, LSTM, dan GRU secara berurutan adalah **-0.63, -0.77, dan -1.09**. Sebaliknya, pada wilayah urban seperti **Metro dan Bandar Lampung**, hasilnya lebih baik dengan nilai masing-masing **0.29, 0.32, dan 0.34 serta 0.29, 0.32, dan 0.35**. Meskipun akurasi model tergolong buruk, data prediksi tetap memiliki klasifikasi yang sesuai dengan data aktual dari API Weatherbit.

5.2 Saran

Berdasarkan Penelitian yang telah dilakukan terdapat beberapa saran dapat dikembangkan terkait dengan penelitian implementasi RNN, LSTM dan GRU untuk memprediksi kualitas udara, antara lain.

1. melakukan peningkatan data dan penambahan fitur agar dapat meningkatkan akurasi.
2. Penggunaan hyperparameter tuning lainnya untuk mencari konfigurasi yang lebih optimal
3. Penggunaan metode hybrid dengan metode statistik seperti *ARIMA*, dsb. kemungkinan memiliki kualitas model yang lebih baik.
4. Penggunaan Berbagai parameter polutan yang terkait dengan kualitas udara memiliki hasil yang beragam di berbagai penelitian terkait kualitas udara. Oleh karena itu, penggunaan seluruh/ sebagian polutan dapat dipertimbangkan dalam penelitian terkait kualitas.

DAFTAR PUSTAKA

- [1] “Dinas Lingkungan Hidup Provinsi Lampung, “Laporan Indeks Kualitas Lingkungan Hidup Provinsi Lampung Tahun 2022,” Lampung: Dinas Lingkungan Hidup Provinsi Lampung, 2022.
- [2] Badan Penelitian dan Pengembangan Kesehatan, “Laporan Riskesdas Lampung Tahun 2018,” Jakarta: Badan Penelitian dan Pengembangan Kesehatan, 2019.
- [3] Kementerian Lingkungan Hidup Republik Indonesia, *Peraturan Menteri Negara Lingkungan Hidup Nomor 12 Tahun 2010 Tentang Pelaksanaan Pengendalian Pencemaran Udara di Daerah*.
- [4] H. Bartelds, H. Heich, and K. Léger, “European Regional Development Fund Regional Initiative Project: Common Information to European Air CAQI Air quality index Comparing Urban Air Quality across Borders-2012,” [Online]. Available: www.citeair.eu
- [5] Kementerian Lingkungan Hidup dan Kehutanan Republik Indonesia, *Peraturan Menteri Lingkungan Hidup dan Kehutanan tentang Indeks Standar Pencemar Udara*, Nomor P.14/MENLHK/SETJEN/KUM.1/7/2020, 2020.
- [6] D. Pengendalian, P. Udara, K. Lingkungan, H. Dan, and K. Jakarta, “Pengembangan Perumusan Indeks Kualitas Udara sebagai Komponen IKLH pada RPJMN 2020-2024,” 2018.
- [7] J. Wang, L. Jin, X. Li, S. He, M. Huang, and H. Wang, “A Hybrid Air Quality Index Prediction Model Based on CNN and Attention Gate Unit,” *IEEE Access*, vol. 10, pp. 113343–113354, 2022, doi: 10.1109/ACCESS.2022.3217242.
- [8] B. K. Hidayatullah, M. Kallista, C. Setianingsih, P. S1, and T. Komputer, “Prediksi Indeks Standar Pencemar Udara Menggunakan Metode Long Short-Term Memory Berbasis Web (Studi Kasus Pada Kota Jakarta),” [Online]. Available: <https://data.jakarta.go.id/>
- [9] A. Khumaidi, R. Raafi, I. Permana Solihin, and J. Rs Fatmawati, “Pengujian Algoritma Long Short Term Memory untuk Prediksi Kualitas Udara dan Suhu Kota Bandung,” *Jurnal Telematika*, vol. 15, no. 1.

- [10] B. Arief, H. Kholifatullah, and A. Prihanto, “Penerapan Metode Long Short Term Memory Untuk Klasifikasi Pada Hate Speech,” *Journal of Informatics and Computer Science*, vol. 04, 2023.
- [11] J. Cahyani, S. Mujahidin, and T. P. Fiqar, “Implementasi Metode Long Short Term Memory (LSTM) untuk Memprediksi Harga Bahan Pokok Nasional,” *Jurnal Sistem dan Teknologi Informasi (JustIN)*, vol. 11, no. 2, p. 346, Jul. 2023, doi: 10.26418/justin.v11i2.57395.
- [12] Namora and J. E. Riwurohi, “Prediction of Water Levels on Peatland using Deep Learning,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 2, pp. 234–239, Apr. 2022, doi: 10.29207/resti.v6i2.3919.
- [13] P. Norvig and S. Russell, *Artificial Intelligence: A Modern Approach*, 4th ed. Boston, MA, USA: Pearson, 2021.
- [14] WHO, *Global Air Quality Guidelines*. Geneva, Switzerland: World Health Organization, 2021.
- [15] U.S. EPA Office of Air Quality Planning, *Technical Assistance Document for the Reporting of Daily Air Quality – the Air Quality Index (AQI)*.
- [16] Badan Pusat Statistik Provinsi Lampung, “Lampung dalam Angka 2023,” Lampung: Badan Pusat Statistik Provinsi Lampung, 2023.
- [17] Doreswamy, K. S. Harishkumar, Y. Km, and I. Gad, “Forecasting Air Pollution Particulate Matter (PM2.5) Using Machine Learning Regression Models,” in *Procedia Computer Science*, Elsevier B.V., 2020, pp. 2057–2066, doi: 10.1016/j.procs.2020.04.221.
- [18] L. Fu, J. Li, and Y. Chen, “An innovative decision making method for air quality monitoring based on big data-assisted artificial intelligence technique,” *Journal of Innovation and Knowledge*, vol. 8, no. 2, Apr. 2023, doi: 10.1016/j.jik.2022.100294.
- [19] S. S. Haykin, *Neural Networks and Learning Machines*. Upper Saddle River, NJ, USA: Prentice Hall/Pearson, (10-24), 2009.
- [20] I. Goodfellow, Y. Bengio, and A. Courville, “Deep Learning.”,(373 – 412).
- [21] N. Srivastava, G. Hinton, A. Krizhevsky, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” 2014.
- [22] M. Schuster and K. K. Paliwal, “Bidirectional Recurrent Neural Networks,” 1997.
- [23] V. Plotnikova, M. Dumas, and F. P. Milani, “Applying the CRISP-DM data mining process in the financial services industry: Elicitation of adaptation requirements,” *Data Knowl Eng*, vol. 139, May 2022, doi: 10.1016/j.datak.2022.102013.

- [24] A. Khumaidi, R. Raafi, I. Permana Solihin, and J. Rs Fatmawati, “Pengujian Algoritma Long Short Term Memory untuk Prediksi Kualitas Udara dan Suhu Kota Bandung,” *Jurnal Telematika*, vol. 15, no. 1.
- [25] R. Janarthanan, P. Partheeban, K. Somasundaram, and P. Navin Elamparithi, “A deep learning approach for prediction of air quality index in a metropolitan city,” *Sustain Cities Soc*, vol. 67, Apr. 2021, doi: 10.1016/j.scs.2021.102720.
- [26] F. Wang, J. Cen, Z. Yu, S. Deng, and G. Zhang, “Research on a hybrid model for cooling load prediction based on wavelet threshold denoising and deep learning: A study in China,” *Energy Reports*, vol. 8, pp. 10950–10962, Nov. 2022, doi: 10.1016/j.egyr.2022.08.237.
- [27] A. Nilsen, “Perbandingan Model RNN, Model LSTM, dan Model GRU dalam Memprediksi Harga Saham-Saham LQ45,” *Jurnal Statistika dan Aplikasinya*, vol. 6, no. 1, 2022.
- [28] K. Sofi, A. S. Sunge, S. R. Riady, and A. Z. Kamalia, “Perbandingan Algoritma Linear Regression, LSTM, dan GRU dalam Memprediksi Harga Saham dengan Model Time Series,” *Seminastika*, vol. 3, no. 1, pp. 39–46, Nov. 2021, doi: 10.47002/seminastika.v3i1.275.
- [29] Y. Karyadi and H. Santoso, “Prediksi Kualitas Udara Dengan Metoda LSTM, Bidirectional LSTM, dan GRU.”
- [30] C. B. Pande *et al.*, “Daily scale air quality index forecasting using bidirectional recurrent neural networks: Case study of Delhi, India,” *Environmental Pollution*, vol. 351, p. 124040, 2024, doi: <https://doi.org/10.1016/j.envpol.2024.124040>.
- [31] Flask, “Welcome to Flask — Flask Documentation (2.3.x),” 2024. [Online]. Available: <https://flask.palletsprojects.com/en/2.3.x/>
- [32] Jeffrey L Elman. Distributed representations, simple recurrent networks, and grammatical structure. *Machine learning*, 7(2-3):195–225, 1991.
- [33] M.I. Jordan. Attractor dynamics and parallelism in a connectionst sequential machine. *In Proceedings of the Cognitive Science Conference*, pages 531— - 546, 1986.