

**IMPLEMENTASI METODE *LINEAR REGRESSION* UNTUK MEMPREDIKSI
KUALITAS UDARA DI KOTA MALANG BERDASARKAN POLUTAN
(2019-2023)**

SKRIPSI

**Oleh:
AYU LITASA
NIM. 200605110088**



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

**IMPLEMENTASI METODE *LINEAR REGRESSION* UNTUK
MEMPREDIKSI KUALITAS UDARA DI KOTA MALANG
BERDASARKAN POLUTAN (2019-2023)**

SKRIPSI

Diajukan Kepada :
Fakultas Sains dan Teknologi
Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang
Untuk Memenuhi Salah Satu Persyaratan Dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:
AYU LITASA
NIM. 200605110088

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024**

HALAMAN PERSETUJUAN

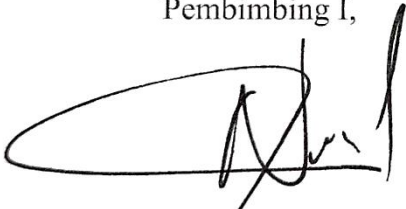
**IMPLEMENTASI METODE *LINEAR REGRESSION* UNTUK
MEMPREDIKSI KUALITAS UDARA DI KOTA MALANG
BERDASARKAN POLUTAN (2019-2023)**

SKRIPSI

**Oleh:
AYU LITASA
NIM. 200605110088**

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 19 September 2024

Pembimbing I,



Dr. Fresy Nugroho, M.T
NIP. 19710722 201101 1 001

Pembimbing II,



Dr. Yunifa Miftachul Arif, M. T
NIP. 19830616 201101 1 004

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Dr. H. Fachrul Kurniawan, M.MT, IPU
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN


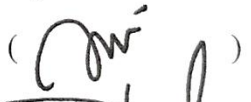


**IMPLEMENTASI METODE *LINEAR REGRESSION* UNTUK
MEMPREDIKSI KUALITAS UDARA DI KOTA MALANG
BERDASARKAN POLUTAN (2019-2023)**

SKRIPSI

Oleh:
AYU LITASA
NIM. 200605110088

Telah Dipertahankan Di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 2 Oktober 2024


Susunan Dewan Penguji

Ketua Penguji	: <u>Hani Nurhayati, M.T</u> NIP. 19780625 200801 2 006	()
Anggota Penguji I	: <u>Ahmad Fahmi Karami, M.Kom</u> NIP. 19870909 202012 1 001	()
Anggota Penguji II	: <u>Dr. Fresy Nugroho, M.T</u> NIP. 19710722 201101 1 001	()
Anggota Penguji III	: <u>Dr. Yunifa Miftachul Arif, M.T</u> NIP. 19830616 201101 1 004	()

Mengetahui,

Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang




Dr. Fachrul Kurniawan, M.MT, IPU
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Ayu Litasa
NIM : 200605110088
Fakultas : Sains dan Teknologi
Program Studi : Teknik Informatika
Judul Skripsi : Implementasi Metode *Linear Regression* Untuk
Memprediksi Kualitas Udara di Kota Malang Berdasarkan
Polutan (2019-2023)

Menyatakan bahwa Skripsi yang saya tulis ini benar-benar hasil karya tulis saya sendiri, bukan merupakan pengambil alihan data, tulisan atau pikiran orang lain yang saya akui hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila di kemudian hari terbukti atau dapat dibuktikan Skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang,
Yang membuat pernyataan,



Ayu Litasa
NIM. 200605110088

HALAMAN MOTTO

“Bisa Kok, Semangat!”

HALAMAN PERSEMBAHAN

*Skripsi ini saya persembahkan untuk kedua orang tua saya, adik, keluarga besar,
sahabat, teman-teman, dan diri saya sendiri.*

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh.

Puji syukur kepada Allah SWT atas limpahan rahmat dan petunjuk-Nya sehingga skripsi dengan judul “Implementasi Metode *Linear Regression* Untuk Memprediksi Kualitas Udara di Kota Malang Berdasarkan Polutan (2019-2023)” dapat terselesaikan. Penulis juga ingin menyampaikan rasa terima kasih kepada:

1. Prof. Dr. M. Zainuddin, M.A., selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Sri Harini, M.Si, selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang.
3. Dr. Ir. Fachrul Kurniawan, M.MT, IPU, selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang
4. Dr. Fresy Nugroho, M.T, selaku dosen pembimbing I dan Dr. Yunifa Miftachul Arif, M.T, selaku dosen pembimbing II atas bimbingan dan arahan kepada penulis dalam penyusunan skripsi ini.
5. Hani Nurhayati, M.T, selaku dosen penguji I dan Ahmad Fahmi Karami, M.Kom, selaku dosen penguji II atas saran yang diberikan sehingga skripsi ini menjadi lebih baik.
6. Ayah Tarwi, S.E, ibu Lilik Sriwahyuni, dan adik Bunga Destiana Anjelita yang selalu mendukung, mendoakan, dan memberi semangat kepada penulis agar dapat menyelesaikan perkuliahan.

7. Seluruh dosen dan staff Program Studi Teknik Informatika UIN Malang atas dukungan selama perkuliahan.
8. Sahabat dan teman-teman yang telah membantu dan memotivasi penulis saat pengerjaan skripsi.
9. Semua pihak yang terlibat saat penulisan baik secara langsung maupun tidak langsung.

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan dan penulis berharap skripsi ini dapat memberikan manfaat bagi pembaca.

Wassalamu alaikum Warahmatullahi Wabarakatuh.

Malang, 20 Oktober 2024

Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
HALAMAN MOTTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xv
ABSTRAK	xv
ABSTRACT.....	xviii
مستخلص البحث	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	6
1.3 Tujuan Penelitian.....	6
1.4 Batasan Masalah	6
1.5 Manfaat Penelitian	7
BAB II STUDI PUSTAKA	8
2.1 Penelitian Terkait	8
2.2 Landasan Teori	10
2.2.1 Polutan	10
2.2.1.1 NO ₂	11
2.2.1.2 SO ₂	11
2.2.1.3 O ₃	11
2.2.1.4 CO.....	12
2.2.1.5 AOD.....	12
2.2.2 <i>Google Earth Engine</i> (GEE).....	13
2.2.3 <i>StandardScaler</i>	16
2.2.4 <i>Linear Regression</i>	17
2.2.5 Evaluasi Metrik.....	18
2.2.5.1 <i>Mean Absolute Error</i> (MAE).....	19
2.2.5.2 <i>Mean Squared Error</i> (MSE).....	19
2.2.5.3 <i>Root Mean Squared Error</i> (RMSE).....	20
BAB III DESAIN PENELITIAN.....	21
3.1 Desain Alur Penelitian.....	21
3.2 Pengumpulan Data Melalui GEE	22
3.3 <i>Preprocessing</i>	24
3.3.1 Penggabungan Data	24
3.3.2 Penanganan <i>Missing Value</i>	24
3.3.3 Normalisasi	26
3.3.4 <i>Split Data</i>	27

3.4 Implementasi Metode.....	28
3.5 Skenario Pengujian.....	30
BAB IV UJI COBA DAN PEMBAHASAN	32
4.1 Pengumpulan Data	32
4.2 <i>Preprocessing</i>	37
4.2.1 Penggabungan Data	37
4.2.2 Penanganan <i>Missing Value</i>	39
4.2.3 Normalisasi	40
4.2.4 <i>Split Data</i>	41
4.3 Implementasi Metode.....	42
4.4 Skenario Pengujian.....	43
4.4.1 Data Lima Tahun (2019-2023)	43
4.4.1.1 Dengan Normalisasi.....	43
4.4.1.1.1 Model 1	44
4.4.1.1.2 Model 2	45
4.4.1.1.3 Model 3	47
4.4.1.2 Tanpa Normalisasi	48
4.4.1.2.1 Model 1	48
4.4.1.2.2 Model 2	49
4.4.1.2.3 Model 3	51
4.4.2 Data Tahunan	52
4.4.2.1 Tahun 2019	52
4.4.2.1.1 Dengan Normalisasi.....	52
4.4.2.1.2 Tanpa Normalisasi.....	56
4.4.2.2 Tahun 2020	61
4.4.2.2.1 Dengan Normalisasi.....	61
4.4.2.2.2 Tanpa Normalisasi.....	65
4.4.2.3. Tahun 2021	69
4.4.2.3.1 Dengan Normalisasi.....	69
4.4.2.3.2 Tanpa Normalisasi.....	74
4.4.2.4 Tahun 2022	78
4.4.2.4.1 Dengan Normalisasi.....	78
4.4.2.4.2 Tanpa Normalisasi.....	83
4.4.2.5 Tahun 2023	88
4.4.2.5.1 Dengan Normalisasi.....	88
4.4.2.5.2 Tanpa Normalisasi.....	92
4.5 Evaluasi.....	96
4.5.1 <i>Mean Absolut Error</i> (MAE).....	96
4.5.2 <i>Mean Squared Error</i> (MSE)	99
4.5.3 <i>Root Mean Squared Error</i> (RMSE)	101
4.6 Integrasi Islam.....	102
4.6.1 Muamalah Mu' Allah	102
4.6.2 Muamalah Mu' Annas	104
4.6.3 Muamalah Mu' Alam	105
BAB V KESIMPULAN DAN SARAN	107
5.1 Kesimpulan	107

5.2 Saran.....	108
----------------	-----

DAFTAR PUSTAKA

DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian.....	21
Gambar 3. 2 Batas Wilayah Kota Malang.....	22
Gambar 3. 3 Boxplot Data Sebelum Normalisasi	26
Gambar 3. 4 Boxplot Data Sesudah Normalisasi.....	27
Gambar 3. 5 <i>Flowchart</i> Metode <i>Linear Regression</i>	29
Gambar 3. 6 Skenario Pengujian.....	30
Gambar 4. 1 Perbandingan Jumlah Baris Data CO.....	33
Gambar 4. 2 Perbandingan Jumlah Baris Data NO ₂	34
Gambar 4. 3 Perbandingan Jumlah Baris Data SO ₂	35
Gambar 4. 4 Perbandingan Jumlah Baris Data O ₃	36
Gambar 4. 5 Perbandingan Jumlah Baris Data AOD.....	36
Gambar 4. 6 Perbandingan Jumlah Baris Data Tiap Parameter Setelah Penghapusan Data Kosong.....	37
Gambar 4. 7 Hubungan Antar Variabel Independen Dengan Variabel Dependen	43
Gambar 4. 8 Perbandingan Hasil Prediksi Model 1 Normalisasi.....	45
Gambar 4. 9 Perbandingan Hasil Prediksi Model 2 Normalisasi.....	46
Gambar 4. 10 Perbandingan Hasil Prediksi Model 3 Normalisasi.....	47
Gambar 4. 11 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi	49
Gambar 4. 12 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi.....	50
Gambar 4. 13 Perbandingan Hasil Prediksi Model 3 Tanpa Normalisasi	51
Gambar 4. 14 Perbandingan Hasil Prediksi Model 1 NormalisasiI 2019	53
Gambar 4. 15 Perbandingan Hasil Prediksi Model 1 Normalisasi 2019.....	54
Gambar 4. 16 Perbandingan Hasil Prediksi Model 3 Normalisasi 2019.....	56
Gambar 4. 17 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi 2019....	57
Gambar 4. 18 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi 2019....	59
Gambar 4. 19 Perbandingan Hasil Evaluasi Model 3 Tanpa Normalisasi 2019 ...	60
Gambar 4. 20 Perbandingan Hasil Prediksi Model 1 Normalisasi 2020.....	62
Gambar 4. 21 Perbandingan Hasil Prediksi Model 2 Normalisasi 2020.....	63
Gambar 4. 22 Perbandingan Hasil Prediksi Model 3 Normalisasi 2020.....	65
Gambar 4. 23 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi 2020	66
Gambar 4. 24 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi	67
Gambar 4. 25 Perbandingan Hasil Prediksi Model 3 Tanpa Normalisasi.....	69
Gambar 4. 26 Perbandingan Hasil Prediksi Model 1 Normalisasi 2021.....	71
Gambar 4. 27 Perbandingan Hasil Prediksi Model 2 Normalisasi 2021.....	72
Gambar 4. 28 Perbandingan Hasil Prediksi Model 3 Normalisasi 2021.....	73
Gambar 4. 29 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi 2021	75
Gambar 4. 30 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi 2021	76
Gambar 4. 31 Perbandingan Hasil Prediksi Model 3 Tanpa Normalisasi 2021	78
Gambar 4. 32 Perbandingan Hasil Prediksi Model 1 Normalisasi 2022.....	80
Gambar 4. 33 Perbandingan Hasil Prediksi Model 2 Normalisasi 2022.....	81
Gambar 4. 34 Perbandingan Hasil Prediksi Model 3 Normalisasi 2022.....	83
Gambar 4. 35 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi 2022	84
Gambar 4. 36 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi 2022	86

Gambar 4. 37 Perbandingan Hasil Prediksi Model 3 Tanpa Normalisasi 2022	87
Gambar 4. 38 Perbandingan Hasil Prediksi Model 1 Normalisasi 2023.....	89
Gambar 4. 39 Perbandingan Hasil Prediksi Model 2 Normalisasi 2023.....	90
Gambar 4. 40 Perbandingan Hasil Prediksi Model 3 Normalisasi 2023.....	91
Gambar 4. 41 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi 2023	93
Gambar 4. 42 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi 2023	94
Gambar 4. 43 Perbandingan Hasil Prediksi Model 3 Tanpa Normalisasi 2023	96

DAFTAR TABEL

Tabel 3. 1 Keterangan Parameter	24
Tabel 3. 2 Data Setelah Penanganan <i>Missing Value</i>	25
Tabel 3. 3 Model <i>Split Data</i>	28
Tabel 4. 1 Data Awal Parameter CO	32
Tabel 4. 2 Penggabungan Data Awal.....	38
Tabel 4. 3 Penggabungan Data Akhir.....	39
Tabel 4. 4 Data Setelah Penanganan <i>Missing Value</i>	40
Tabel 4. 5 Data Setelah Proses Normalisasi.....	40
Tabel 4. 6 Model Split Data Lima Tahun	41
Tabel 4. 7 Model Split Data Pertahun	41
Tabel 4. 8 Hubungan Antar Variabel	43
Tabel 4. 9 Perbandingan Hasil Prediksi Model 1 Normalisasi.....	44
Tabel 4. 10 Hasil Evaluasi Model 1 Normalisasi.....	45
Tabel 4. 11 Perbandingan Hasil Prediksi Model 2 Normalisasi.....	46
Tabel 4. 12 Hasil Evaluasi Model 2 Normalisasi.....	46
Tabel 4. 13 Perbandingan Hasil Prediksi Model 3 Normalisasi.....	47
Tabel 4. 14 Hasil Evaluasi Model 3 Normalisasi	48
Tabel 4. 15 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi.....	48
Tabel 4. 16 Hasil Evaluasi Model 1 Tanpa Normalisasi	49
Tabel 4. 17 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi	49
Tabel 4. 18 Hasil Evaluasi Model 2 Tanpa Normalisasi	50
Tabel 4. 19 Perbandingan Hasil Prediksi Model 3 Tanpa Normalisasi.....	51
Tabel 4. 20 Hasil Evaluasi Model 3 Tanpa Normalisasi	52
Tabel 4. 21 Perbandingan Hasil Prediksi Model 1 Normalisasi 2019.....	53
Tabel 4. 22 Hasil Evaluasi Model 1 Normalisasi 2019.....	53
Tabel 4. 23 Perbandingan Hasil Prediksi Model 2 Normalisasi 2019.....	54
Tabel 4. 24 Hasil Evaluasi Model 1 Normalisasi 2019.....	55
Tabel 4. 25 Perbandingan Hasil Prediksi Model 3 Normalisasi 2019.....	55
Tabel 4. 26 Hasil Evaluasi Model 3 Normalisasi 2019.....	56
Tabel 4. 27 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi	57
Tabel 4. 28 Hasil Evaluasi Model 1 Tanpa Normalisasi 2019	58
Tabel 4. 29 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi 2019	58
Tabel 4. 30 Hasil Evaluasi Model 2 Tanpa Normalisasi	59
Tabel 4. 31 Perbandingan Hasil Prediksi Model 3 Tanpa Normalisasi 2019.....	60
Tabel 4. 32 Hasil Evaluasi Model 3 Tanpa Normalisasi 2019	61
Tabel 4. 33 Perbandingan Hasil Prediksi Model 1 Normalisasi 2020.....	62
Tabel 4. 34 Hasil Evaluasi Model 1 Normalisasi 2020.....	62
Tabel 4. 35 Perbandingan Hasil Prediksi Model 2 Normalisasi 2020.....	63
Tabel 4. 36 Hasil Evaluasi Model 2 Normalisasi 2020.....	64
Tabel 4. 37 Perbandingan Hasil Prediksi Model 3 Normalisasi 2020.....	64
Tabel 4. 38 Perbandingan Hasil Prediksi Model 3 Normalisasi 2020.....	65
Tabel 4. 39 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi 2020.....	66
Tabel 4. 40 Hasil Evaluasi Model 1 Tanpa Normalisasi 2020	66

Tabel 4. 41 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi 2020	67
Tabel 4. 42 Hasil Evaluasi Model 2 Tanpa Normalisasi 2020	68
Tabel 4. 43 Perbandingan Hasil Prediksi Model 3 Tanpa Normalisasi 2020	68
Tabel 4. 44 Hasil Evaluasi Model 3 Tanpa Normalisasi 2020	69
Tabel 4. 45 Perbandingan Hasil Prediksi Model 1 Normalisasi 2021	70
Tabel 4. 46 Hasil Evaluasi Model 1 Normalisasi 2021	71
Tabel 4. 47 Perbandingan Hasil Prediksi Model 2 Normalisasi 2021	72
Tabel 4. 48 Hasil Evaluasi Model 2 Normalisasi 2021	72
Tabel 4. 49 Perbandingan Hasil Prediksi Model 3 Normalisasi 2021	73
Tabel 4. 50 Hasil Evaluasi Model 3 Normalisasi 2021	74
Tabel 4. 51 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi 2021	74
Tabel 4. 52 Hasil Evaluasi Model 1 Tanpa Normalisasi 2021	75
Tabel 4. 53 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi 2021	76
Tabel 4. 54 Hasil Evaluasi Model 2 Tanpa Normalisasi 2021	76
Tabel 4. 55 Perbandingan Hasil Prediksi Model 3 Tanpa Normalisasi 2021	77
Tabel 4. 56 Hasil Evaluasi Model 3 Tanpa Normalisasi 2021	78
Tabel 4. 57 Perbandingan Hasil Prediksi Model 1 Normalisasi 2022	79
Tabel 4. 58 Hasil Evaluasi Model 1 Normalisasi 2022	80
Tabel 4. 59 Perbandingan Hasil Prediksi Model 2 Normalisasi 2022	81
Tabel 4. 60 Hasil Evaluasi Model 2 Normalisasi 2022	81
Tabel 4. 61 Perbandingan Hasil Prediksi Model 3 Normalisasi 2022	82
Tabel 4. 62 Hasil Evaluasi Model 3 Normalisasi 2022	83
Tabel 4. 63 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi 2022	84
Tabel 4. 64 Hasil Evaluasi Model 1 Tanpa Normalisasi 2022	85
Tabel 4. 65 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi 2022	85
Tabel 4. 66 Hasil Evaluasi Model 2 Tanpa Normalisasi 2022	86
Tabel 4. 67 Perbandingan Hasil Prediksi Model 3 Tanpa Normalisasi 2022	87
Tabel 4. 68 Hasil Evaluasi Model 3 Tanpa Normalisasi 2022	87
Tabel 4. 69 Perbandingan Hasil Prediksi Model 1 Normalisasi 2023	88
Tabel 4. 70 Hasil Evaluasi Model 1 Normalisasi 2023	89
Tabel 4. 71 Perbandingan Hasil Prediksi Model 2 Normalisasi 2023	90
Tabel 4. 72 Hasil Evaluasi Model 2 Normalisasi 2023	90
Tabel 4. 73 Perbandingan Hasil Prediksi Model 3 Normalisasi 2023	91
Tabel 4. 74 Hasil Evaluasi Model 3 Normalisasi 2023	92
Tabel 4. 75 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi 2023	93
Tabel 4. 76 Hasil Evaluasi Model 1 Tanpa Normalisasi 2023	93
Tabel 4. 77 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi 2023	94
Tabel 4. 78 Hasil Evaluasi Model 2 Tanpa Normalisasi 2023	95
Tabel 4. 79 Perbandingan Hasil Prediksi Model 3 Tanpa Normalisasi 2023	95
Tabel 4. 80 Hasil Evaluasi Model 3 Tanpa Normalisasi 2023	96
Tabel 4. 81 MAE Lima Tahun	97
Tabel 4. 82 MAE Pertahun	98
Tabel 4. 83 MSE Data Lima Tahun	99
Tabel 4. 84 MSE Data Pertahun	100
Tabel 4. 85 RMSE Data Lima Tahun	101
Tabel 4. 86 RMSE Data Pertahun	102

ABSTRAK

Litasa, Ayu. 2024. **Implementasi Metode *Linear Regression* Untuk Memprediksi Kualitas Udara di Kota Malang Berdasarkan Polutan (2019-2023)**. Skripsi, Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Fresy Nugroho, M.T. (II) Dr. Yunifa Miftachul Arif, M.T.

Kata Kunci: *Linear Regression*, Kualitas Udara, Polutan.

Manusia akan selalu berinteraksi dengan komponen-komponen penyusun lingkungan, seperti udara. Kualitas udara yang baik akan menghasilkan kehidupan yang sehat. Akan tetapi, seiring dengan perubahan aktivitas manusia seperti urbanisasi, pembangunan, dan industrialisasi menyebabkan penambahan polutan pencemar udara sehingga menurunkan kualitas udara. Hal tersebut banyak terjadi terutama di kota-kota besar, seperti kota Malang. Penurunan kualitas udara ini memberi dampak buruk mulai dari permasalahan kesehatan hingga memicu perubahan iklim. Untuk mengantisipasi dampak buruk tersebut, penelitian ini mengimplementasikan metode *linear regression* untuk memprediksi kualitas udara di Kota Malang berdasarkan polutan. Polutan yang digunakan terdiri dari CO, NO₂, SO₂, O₃, PM_{2.5}. Dalam penelitian ini, data dianalisis menggunakan data lima tahunan (2019-2023) dan data pertahun mulai dari 2019 hingga 2023. Masing-masing data dibagi menjadi dua skenario yaitu dengan normalisasi dan tanpa normalisasi. Masing-masing skenario terdapat tiga model *split data* yaitu 75:25, 80:20, dan 90:10. Setiap model dievaluasi menggunakan *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Hasil evaluasi menunjukkan bahwa dengan atau tanpa normalisasi data hanya memiliki pengaruh yang sangat kecil pada hasil evaluasi dan pengujian dengan data lima tahunan lebih sesuai untuk penelitian ini. Selain itu, semakin banyak selisih antara hasil prediksi dan nilai aktual membuat hasil evaluasi MAE, MSE, dan RMSE menjadi lebih tinggi.

ABSTRACT

Litasa, Ayu. 2024. **Implementation of the Linear Regression Method to Predict Air Quality in Malang City Based on Pollutants (2019-2023)**. Thesis, Informatics Engineering Study Program, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University, Malang. Supervisor: (I) Dr. Fresy Nugroho, M.T. (II) Dr. Yunifa Miftachul Arif, M.T.

Humans constantly interact with their environment components, such as air. Good air quality will lead to a healthy life. However, the changes in human activities, such as urbanization, development, and industrialization, increase air pollutants and decrease air quality. It often occurs in big cities like Malang City. The air quality brings various bad impacts, from causing health issues to triggering climate change. To anticipate them, the research implements a regression linear method to predict air quality in Malang city based on pollutants. The researcher employed pollutants consisting of CO, NO₂, SO₂, O₃, and PM_{2.5}. In this study, data was analyzed using five-year data (2019-2023) and annual data from 2019 to 2023. Each data was split into two scenarios: using and without normalization. Each scenario had three split data models: 75:25, 80:20, and 90:10. Each model was evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The evaluation result shows that using or without data normalization has an insignificant influence on the result. The test using five-year data is more suitable for the research. Furthermore, the higher the prediction result and actual score difference, the higher the evaluation result of MAE, MSE, and RMSE.

Keywords: Linear Regression, Air Quality, Pollutants

مستخلص البحث

لبناسا، أيو. 2024. تنفيذ طريقة الانحدار الخطي للتنبؤ بجودة الهواء في مدينة مالانج بناء على الملوثات (2019-2023). البحث الجامعي، قسم الهندسة المعلوماتية، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول: د. فريسي نوجروهو، الماجستير. المشرف الثاني: د. يونيفة مفتاح العارف، الماجستير.

الكلمات الرئيسية: انحدار خطي، جودة الهواء، الملوثات.

ستفاعل البشر دائما مع المكونات التي تشكل البيئة، مثل الهواء. جودة الهواء الجيدة ستؤدي إلى حياة صحية. ومع ذلك، إلى جانب التغيرات في الأنشطة البشرية مثل التحضر والتنمية والتصنيع، فإنه يتسبب في إضافة ملوثات الهواء، وبالتالي انخفاض جودة الهواء. يحدث هذا كثيرا، خاصة في المدن الكبرى، مثل مدينة مالانج. هذا الانخفاض في جودة الهواء له تأثير سيء يتراوح من المشاكل الصحية إلى التسبب في تغير المناخ. لتوقع هذه الآثار الضارة، نفذ هذا البحث طريقة الانحدار الخطي للتنبؤ بجودة الهواء في مدينة مالانج بناء على الملوثات. تتكون الملوثات المستخدمة من CO و NO₂ و SO₂ و PM_{2.5} و O₃. في هذا البحث، تم تحليل البيانات باستخدام بيانات خمس سنوات متوالية (2019-2023) والبيانات السنوية من 2019 إلى 2023. تنقسم كل بيانات إلى سيناريوهين، وهما التطبيع وعدم التطبيع. لكل سيناريو، هناك ثلاثة نماذج لتقسيم البيانات، وهي 75:25 و 80:20 و 90:10. تم تقييم كل نموذج باستخدام متوسط الخطأ المطلق (MAE) ومتوسط الخطأ التربيعي (MSE) وجذر متوسط الخطأ التربيعي (RMSE). أظهرت نتائج التقييم أنه مع أو بدون تطبيع البيانات لم يكن له سوى تأثير ضئيل جدا على نتائج التقييم والاختبار مع بيانات خمس سنوات متوالية كان أكثر ملاءمة لهذا البحث. بالإضافة إلى ذلك، كلما زاد الاختلاف بين النتائج المتوقعة والقيمة الفعلية، زادت نتائج تقييم MAE و MSE و RMSE

BAB I

PENDAHULUAN

1.1 Latar Belakang

Manusia hidup di dalam sebuah lingkungan dan melakukan interaksi dengan komponen-komponen penyusun lingkungan. Salah satu komponen penyusun lingkungan adalah udara. Udara sangat penting untuk menjaga kelangsungan kehidupan makhluk hidup. Kualitas udara yang baik akan menghasilkan kehidupan yang sehat. Tetapi seiring dengan perubahan aktivitas manusia seperti urbanisasi, pembangunan, dan industrialisasi menyebabkan bertambahnya polutan pencemar udara sehingga menurunkan kualitas udara. Penurunan kualitas udara ini menimbulkan berbagai penyakit bagi manusia seperti, iritasi mata, gangguan infeksi saluran pernapasan atas (ISPA) hingga kanker paru-paru (Dian et al., 2023). Selain itu, penurunan kualitas udara juga menyebabkan peningkatan suhu permukaan bumi yang akan memicu perubahan iklim. Perubahan iklim mempunyai dampak serius terhadap ekosistem dan ekonomi. Dampak terhadap ekosistem berupa punahnya beberapa spesies hewan atau tumbuhan yang mengakibatkan berkurangnya keanekaragaman hayati. Pembangunan ekonomi, perubahan cuaca ekstrim dapat menyebabkan kekeringan dan banjir sehingga akan mempengaruhi infrastruktur bangunan, tata letak pemukiman, pembukaan lahan, dan mempengaruhi produksi pertanian (Maimaiti, 2023).

Indonesia selama dua dekade terakhir telah mengalami penurunan kualitas udara karena meningkatnya partikulat PM 2.5 dari $8 \mu\text{g}/\text{m}^3$ hingga $22 \mu\text{g}/\text{m}^3$

(Greenstone & Fan, 2019). Pulau Jawa menjadi pulau yang tergolong memiliki kualitas udara yang kurang sehat diantara pulau-pulau lainnya. Penurunan kualitas udara ini tentu menjadi perhatian, terutama kota-kota besar yang memiliki jumlah penduduk yang banyak. Lingkungan yang padat penduduk akan menyebabkan penurunan kualitas udara. Hal ini dikarenakan dilakukannya aktivitas seperti merokok, penggunaan transportasi, pembakaran lahan, dan kegiatan industri. (Astriyani et al., 2022).

Salah satu kota besar di Pulau Jawa adalah Kota Malang. Kota Malang merupakan kota terbesar kedua di Provinsi Jawa Timur. Kota ini dikenal sebagai kota pendidikan dan pariwisata sehingga menjadi tujuan untuk belajar dan bekerja yang mengakibatkan penambahan jumlah penduduk. Penambahan ini berbanding lurus dengan meningkatnya aktivitas dan mobilitas. Sehingga terjadi kenaikan jumlah kendaraan bermotor. Selain itu, ada banyak jenis industri di Kota Malang, termasuk usaha kecil, menengah, besar, bahkan rumah-rumahan. Badan Pusat Statistik (BPS) Kota Malang melaporkan terdapat 5.920 industri di kota ini pada tahun 2021, dan jumlahnya naik menjadi 6.722 pada tahun 2022 (Martha, 2022). Tentu saja hal yang menyebabkan turunnya kualitas udara terjadi di Kota Malang.

Untuk mencegah penurunan kualitas udara diperlukan sebuah pengetahuan tentang penyebab dan kandungan polutan yang dapat mempengaruhinya. Kementerian Lingkungan Hidup dan Kehutanan dalam PP nomor 14 tahun 2020, untuk mengetahui indeks standar pencemaran udara dapat menggunakan parameter partikulat (PM_{10}), partikulat ($PM_{2.5}$), karbon monoksida (CO), nitrogen dioksida (NO_2), sulfur dioksida (SO_2), ozon (O_3), dan hidrokarbon

(HC). Selain itu, polutan yang dapat ditemui di kawasan industri adalah karbon monoksida (CO), nitrogen dioksida (NO₂), sulfur dioksida (SO₂), hidrokarbon (HC) (Safira et al., 2022).

Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) memberikan informasi tentang kualitas udara. Berdasarkan Peraturan BMKG nomor 2 tahun 2020, BMKG memiliki data tentang konsentrasi parameter partikulat (PM₁₀, PM_{2.5}), sulfur dioksida (SO₂), nitrogen dioksida (NO₂), ozon (O₃), karbon monoksida (CO), karbon dioksida (CO₂), dan metan (CH₄). Akan tetapi, tidak semua stasiun pemantauan memiliki parameter dengan lengkap.

Untuk mengetahui konsentrasi polutan dapat menggunakan pengindraan jauh dengan bantuan satelit. Sentinel-5 Copernicus merupakan satelit yang memiliki fokus tentang pengendalian polusi udara dan interaksi iklim. Sentinel memiliki produk berupa data sulfur dioksida (SO₂), nitrogen oksida (NO₂), ozon (O₃), karbon monoksida (CO), dan metan (CH₄). Data tersebut berupa gambar real-time, gambar offline, dan gambar pemrosesan ulang. Sentinel-5 dapat diakses melalui *Google Earth Engine* (GEE) dengan menggunakan bahasa JavaScript. Terdapat korelasi kuat terhadap CO, SO₂, NO₂, dan O₃ yang diambil dari sentinel-5 dan stasiun polusi udara, sehingga sentinel-5 berdasarkan GEE dapat menjadi pendekatan yang efisien untuk memantau konsentrasi polutan (Garajeh et al., 2023). Selain menggunakan Sentinel-5 juga dapat menggunakan satelit *Moderate Resolution Imaging Spectroradiometer* (MODIS). Satelit ini memiliki bands *Aerosol Optical Depth* (AOD) yang digunakan untuk pengukuran terhadap penghamburan dan penyerapan jumlah sinar oleh partikel udara. Sebagian besar studi tentang AOD

dan $PM_{2.5}$ membuktikan bahwa AOD dan $PM_{2.5}$ memiliki hubungan *linear*. Nilai AOD dapat digunakan untuk memprediksi konsentrasi $PM_{2.5}$ (Nur et al., 2023).

Prediksi dapat dilakukan dengan menggunakan beberapa teknik. Contoh metode populer digunakan untuk memprediksi adalah *linear regression*. *Linear regression* secara umum digunakan untuk mengidentifikasi hubungan antar variabel sehingga dapat memudahkan untuk melakukan prediksi. Selain itu, metode ini juga telah digunakan pada beberapa penelitian tentang udara. *Linear regression* dapat digunakan untuk mengetahui korelasi antara data polutan yang didapatkan dari GEE dan data polutan dari Stasiun polusi udara memiliki nilai *Root Mean Square Error* setiap variabel berada di rentang 0,13 sampai 4,62 (Garajeh et al., 2023). Metode *Linear regression* juga digunakan untuk memprediksi kualitas udara di dunia menunjukkan bahwa *linear regression* lebih unggul dibandingkan *random forest regression* (Anggraini et al., 2024).

Metode *linear regression* sebagai model prediksi dapat membantu untuk mempelajari dan mengerti tentang perubahan kualitas udara pada waktu yang akan datang. Metode ini juga dapat menjadi sarana bagi manusia untuk mengetahui tanda-tanda kebesaran Allah melalui udara. Dalam Al-Qur'an surah Al-Jasyiah ayat 5, Allah berfirman:

وَإِخْتِلَافِ اللَّيْلِ وَالنَّهَارِ وَمَا أَنْزَلَ اللَّهُ مِنَ السَّمَاءِ مِنْ رِزْقٍ فَأَحْيَا بِهِ الْأَرْضَ بَعْدَ مَوْتِهَا وَتَصْرِيْفِ الرِّيحِ آيَاتٍ لِّقَوْمٍ يَعْقِلُونَ

“ dan pada pergantian malam dan siang dan hujan yang diturunkan Allah dari langit lalu dengan (air hujan) itu dihidupkan-Nya bumi setelah mati (kering); dan

pada perkisaran angin terdapat pula tanda-tanda (kebesaran Allah) bagi kaum yang mengerti.” (Q.S Al-Jasyiah 45:5).

Ayat ini menjadi pembelajaran kepada kita semua agar senantiasa belajar sehingga dapat menyadari dan mengerti tentang kebesaran Allah. Seperti halnya dengan penggunaan metode *linear regression* untuk memprediksi kualitas udara sebagai salah satu bentuk cara berpikir guna menyadari dan mengerti tentang kebesaran Allah melalui ilmu pengetahuan. Hal ini sejalan dengan Al-Qur’an surat An-Nahl ayat 11:

يُثْبِتُ لَكُمْ بِهِ الزَّرْعَ وَالزَّيْتُونَ وَالنَّخِيلَ وَالْأَعْنَابَ وَمِنْ كُلِّ الثَّمَرَاتِ إِنَّ فِي ذَلِكَ لَآيَةً لِّقَوْمٍ يَتَفَكَّرُونَ

“Dengan (air hujan) itu Dia menumbuhkan untukmu tumbuh-tumbuhan, zaitun, kurma, anggur, dan segala macam buah-buahan. Sesungguhnya pada yang demikian itu benar-benar terdapat tanda (kebesaran Allah) bagi orang yang berpikir.” (Q.S. Al-Nahl 16:11).

Selain itu, seperti yang telah diuraikan pada paragraf awal. Manusia merupakan kontributor utama dari menurunnya kualitas udara. Sebagai makhluk yang dilimpahi ilmu pengetahuan dan akal untuk berpikir, terkadang hal tersebut justru sering mendorong manusia untuk berbuat kerusakan. Walaupun manusia dilimpahi ilmu pengetahuan yang tinggi, tetapi manusia juga menjadi penyebab kerusakan muka bumi sehingga Allah akan menarik nur dalam hatinya dan tidak akan datang pertolongan dan perlindungan dari-Nya (Miskhahuddin, 2019). Oleh sebab itu, manusia diharapkan dapat terus berbuat baik dan turut serta menjaga kualitas udara agar dampak-dampak buruk tersebut dapat dihindari. Hal ini sesuai dengan firman Allah dalam Al- Qur’an Surah al- ‘Araf ayat ke 56.

وَلَا تُفْسِدُوا فِي الْأَرْضِ بَعْدَ إِصْلَاحِهَا وَادْعُوهُ خَوْفًا وَطَمَعًا إِنَّ رَحْمَتَ اللَّهِ قَرِيبٌ مِّنَ الْمُحْسِنِينَ

“Dan janganlah kamu membuat kerusakan di muka bumi, sesudah (Allah) memperbaikinya dan Berdoalah kepada-Nya dengan rasa takut (tidak akan diterima) dan harapan (akan dikabulkan). Sesungguhnya rahmat Allah Amat dekat kepada orang-orang yang berbuat baik.” (Q.S. Al- ‘Araf 7:56).

Berdasarkan latar belakang yang telah dipaparkan, penelitian ini akan mengimplementasikan metode *linear regression* untuk melakukan prediksi kualitas udara. Sehingga penelitian ini berjudul “**Implementasi Metode *Linear Regression* Untuk Memprediksi Kualitas Udara di Kota Malang Berdasarkan Polutan (2019-2023)**”.

1.2 Rumusan Masalah

Bagaimana memprediksi kualitas udara di Kota Malang berdasarkan polutan menggunakan metode *linear regression*?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah mengimplementasikan metode *linear regression* untuk memprediksi kualitas udara di Kota Malang berdasarkan polutan.

1.4 Batasan Masalah

Data yang digunakan adalah data polutan dengan parameter NO₂, SO₂, CO, dan O₃ yang didapatkan dari Sentinel-5 dan parameter AOD yang didapatkan dari Modis melalui *Google Earth Engine* pada wilayah Kota Malang dalam rentang waktu dari 2019 hingga 2023.

1.5 Manfaat Penelitian

Berdasarkan rumusan masalah, manfaat yang diharapkan dari penelitian ini dapat menambah wawasan tentang kualitas udara dan menjadi referensi untuk penelitian selanjutnya dalam menentukan kualitas udara kota Malang.

BAB II

STUDI PUSTAKA

2.1 Penelitian Terkait

Penelitian tentang kualitas udara dunia telah dilakukan oleh Anggraini et al., (2024), peneliti mengimplementasikan empat metode yang berbeda yaitu, SLR, MLR, SRFR, dan MRFR. *Multi metode* seperti MLR dan MRFR dianggap lebih unggul daripada metode Tunggal seperti SLR dan RFR jika berdasarkan nilai R2 dan nilai RMSE. Tetapi, jika secara keseluruhan, metode MLR dianggap paling baik. Penelitian ini menggunakan variabel polutan seperti NO₂, CO, SO₂, dan AOD yang berasal dari Sentinel-5P dan MODIS.

Latief et al., (2022) menggunakan beberapa rumus turunan dari metode *linear regression* untuk memprediksikan indeks standar pencemar udara Jakarta. Penelitian ini menggunakan dataset yang didapatkan dari *website* Opendata Jakarta. Parameter yang digunakan yaitu *factor, critical, max, category*, PM₁₀, PM_{2.5}, SO₂, CO, O₃, CO₂, tanggal, dan stasiun. Metode regresi yang cocok diterapkan pada dataset ini adalah *Multi Linear* dan Ridge. Hasil *metrics* dari kedua metode tersebut nilai validasi dan testnya tidak berbeda jauh sedangkan metode *polynomial* hasil nilai validasi dan testnya memiliki perbedaan yang jauh.

Pada penelitian yang dilakukan oleh Garajeh et al., (2023) menggunakan citra sentinel-5p yang berdasarkan GEE untuk mengambil parameter CO, NO₂, SO₂, dan O₃. Selain itu juga menggunakan data dengan parameter yang sama pada stasiun polusi udara. Kedua data dicari korelasi dengan menggunakan Metode

Linear Regression didapatkan bahwa terdapat korelasi yang kuat antara Sentinel-5p dengan stasiun polusi udara.

Kumar & Jamatia, (2023) Pada penelitian yang dilakukan oleh Kumar & Jamatia, (2023) yang berjudul “*Prediction of Air Quality Using Machine Learning*” meneliti kualitas udara di kota Bangalore, India. Penelitian ini mengimplementasikan tujuh metode *machine learning*, yaitu *Linear Regression*, RR, Lasso R, DTR, RFR, XGBoost, dan ANN. Hasil dari kesimpulan ini didapatkan bahwa kualitas udara hanya dapat diprediksi dalam jangka pendek. Jika dilakukan melebihi 2 jam tidak akan memberikan hasil yang baik karena interval waktu antara prediksi dan hasil pelatihan terlalu lama.

Pebralia, (2022) menggunakan *multi linear regression* untuk memprediksi curah hujan. Data yang digunakan merupakan data yang berasal dari Kaggle dan berjumlah 6.574 data. Parameter yang digunakan yaitu WIND, T.MAX, T.MIN sebagai variabel bebas dan RAIN sebagai variabel terikat. Hasil performanya sebagai berikut MAE sebesar 2.23, MSE sebesar 14.02, dan RMSE sebesar 3.74.

Penelitian yang dilakukan Baskara et al., (2023) menggunakan *Multiple Linear Regression* untuk memprediksi penjualan *Crude Palm Oil* (CPO). Penelitian ini menggunakan variabel independen produksi CPO, jumlah minyak yang disimpan ditangki 1 dan jumlah minyak yang disimpan ditangki 2 sedangkan variabel dependennya adalah pengiriman CPO. Hasil dari penelitian ini didapatkan bahwa nilai koefisien determinasi (*R-Squared*) sebesar 0,9546 dengan artian bahwa 95,46% variabelitas CPO dapat dijelaskan oleh variabel dependen. Hasil uji MAPE

dan RMSE juga menunjukkan tingkat keakuratan yang baik dalam memprediksi penjualan CPO.

Multiple linear regression juga digunakan oleh Alita et al., (2021) untuk memprediksikan performa pegawai yang cocok untuk dipromosikan. Data yang digunakan berupa *employee performance* sebagai variabel dependen dan variabel independen seperti *orientation, integrity, discipline, cooperation, dan performance*. Dasi penelitian ini didapatkan MAE sebesar 2,24% dan RMSE 3,88%. Dari hasil studi tersebut dapat disimpulkan bahwa *multiple linear regression* memiliki performa yang bagus dengan nilai *error* yang kecil sehingga dapat menghasilkan prediksi yang baik.

2.2 Landasan Teori

2.2.1 Polutan

Polutan ialah zat pencemar yang dapat menyebabkan pencemaran terhadap lingkungan. Jika suatu zat melebihi ambang batas atau batas normal maka dapat dikategorikan sebagai zat pencemar. Terdapat tiga jenis polutan jika didasarkan pada wujudnya, yaitu padat, cair, dan gas. Polutan padat misalnya kaleng, besi, logam, plastic, dan lain-lain. Polutan cair misalnya detergen, tumpahan minyak, pestisida, dan lainnnya. Polutan gas seperti NO₂, O₃, CO, CFC, dan metana. (Lumeno et al., 2016).

Berdasarkan jenisnya, polutan udara dibagi menjadi polutan primer dan polutan sekunder. Polutan primer dihasilkan langsung oleh sumber tertentu dan 90% dari jumlah polutan yang ada di udara merupakan polutan primer. Polutan primer berasal dari polutan gas (NO₂, SO₂, CO, CO₂, HC, metana) dan partikulat

atau partikel (zat padat dan cair yang dihasilkan secara mekanik dan pembakaran) sedangkan polutan sekunder terbentuk dari reaksi dua atau lebih pencemar primer di atmosfer. Polutan sekunder dipengaruhi oleh kondisi iklim topografi lokal, derajat fotoaktivasi, dan konsentrasi *relative* reaktan (Husna et al., 2022).

2.2.1.1 NO₂

Nitrogen dioksida (NO₂) ialah gas polutan yang dihasilkan dari kegiatan seperti pembakaran bahan bakar kendaraan bermotor, kegiatan industri, dan pembakaran sampah. Gas polutan ini memiliki karakteristik berbau tajam dan berwarna coklat kemerah-merahan (Maherdyta et al., 2022). NO₂ dapat mengakibatkan dampak buruk bagi kesehatan manusia, seperti rusaknya saluran pernapasan, iritasi mata, hingga kerusakan jantung.

2.2.1.2 SO₂

Sulfur dioksida (SO₂) ialah gas polutan yang dihasilkan dari pembakaran bahan bakar yang mengandung belerang. Gas Polutan ini memiliki karakteristik berbau tajam dan tidak berwarna. Jika SO₂ bereaksi dengan uap air di udara bebas akan berubah menjadi senyawa asam sulfat (H₂SO₄). Saat senyawa ini bereaksi dengan air hujan akan mengakibatkan hujan asam. Hujan asam memiliki dampak yang buruk bagi lingkungan dan makhluk hidup (Masito, 2018).

2.2.1.3 O₃

Ozon (O₃) ialah gas polutan yang dapat dihasilkan secara alami dan reaksi kimia. Secara alami ozon berasal dari hidrokarbon yang dilepaskan oleh tanaman, tanah serta ozon yang turun dari lapisan stratosfer. Secara reaksi kimia, ozon

terbentuk dari prekursor NO_x , VOC_x , SH_4 , dan CO saat mendapatkan radiasi dari matahari (Hikmah et al., 2022). Ozon memiliki karakteristik tidak berwarna tetapi menjadi komponen utama kabut asap. Gas ini dapat meningkatkan resiko penyakit pernapasan seperti asma, iritasi paru-paru, dan batuk.

2.2.1.4 CO

Karbon monoksida (CO) ialah gas polutan yang dihasilkan dari emisi kendaraan bermotor. Gas polutan ini memiliki karakteristik tidak berbau tetapi beracun. CO memiliki kemampuan untuk mengikat hemoglobin darah sehingga kapasitas darah untuk mengikat oksigen akan menurun. Sehingga dapat mengganggu kinerja otak (Raming et al., 2022).

2.2.1.5 AOD

Aerosol Optical Depth (AOD) ialah nilai hamburan radiasi matahari ke permukaan bumi yang dapat mewakili jumlah aerosol di atmosfer. Nilai AOD dapat dipengaruhi oleh pembakaran fosil dan biomassa yang menjadi sumber utama partikel halus sehingga menyebabkan polusi pada atmosfer. Selain itu juga dapat dipengaruhi oleh perubahan iklim (Lewar & Darmawan, 2022).

Sebagian besar studi tentang AOD dan $\text{PM}_{2.5}$ membuktikan bahwa AOD dan $\text{PM}_{2.5}$ memiliki hubungan *linear*. Nilai AOD dapat digunakan untuk memprediksi konsentrasi $\text{PM}_{2.5}$ (Nur et al., 2023). $\text{PM}_{2.5}$ merupakan polutan utama yang sangat mempengaruhi polusi udara. $\text{PM}_{2.5}$ berasal dari sumber yang sama dengan NO_2 , SO_2 , O_3 , dan CO. Menurut IQAir, (2022), $\text{PM}_{2.5}$ memiliki komponen kimia yang

berasal dari reaksi kimia gas seperti SO₂, NO₂, ammonia, karbon hitam, debu mineral, air, dan senyawa organik yang mudah menguap.

2.2.2 Google Earth Engine (GEE)

Google Earth Engine (GEE) ialah *platform* dengan basis *cloud* yang digunakan untuk analisis data geospasial global dan dapat diakses melalui *web browser*. Platform yang dirilis tahun 2010 ini mendukung banyak jenis data citra geospasial seperti Sentinel, Landsat, dan MODIS yang dapat diakses di satu tempat tanpa harus melalui *platform* USGS dan ESA. Data yang ada pada GEE juga tersedia secara *real time* sehingga dapat dilakukan pemrosesan secara cepat dan terkini terkait berbagai isu lokal maupun global. Selain itu, GEE juga memiliki data observasi berskala *petabyte* yang dapat diunduh secara publik dan tidak berbayar (Sukoco, 2022).

Katalog GEE memiliki berbagai jenis data yang dapat dibagi menjadi kategori tertentu. Menurut Ramdani, (2021) katalog GEE dapat dibagi menjadi empat, yaitu berdasarkan jenis data, sifat, area observasi, dan waktu. Berdasarkan jenisnya, data dapat dikelompokkan menjadi dua yaitu:

a. Data Primer

Data primer merupakan data mentah yang dapat diolah sesuai dengan kebutuhan penelitian. Selain itu, data primer juga merupakan data yang mendominasi di katalog GEE. Contoh dari data primer antara lain, data MODIS, data Sentinel 1-5, data Landsat 1-8, dan ASTER.

b. Data Sekunder

Data sekunder merupakan data yang telah diproses dan diolah menjadi nilai tertentu. Data ini dihasilkan dari model atau analisis dari berbagai data yang berbeda. Contoh dari data sekunder adalah data tentang kependudukan, seperti kepadatan penduduk (*Global Human Settlement Layers*), data kondisi tanah (*OpenLandMap*), dan data penggunaan lahan (*Copernicus Global Land Cover Layer*).

Berdasarkan sifatnya, data dapat dikelompokkan menjadi dua, yaitu:

a. Data Satelit Pasif

Data pasif merupakan data yang bergantung pada pancaran energi matahari. Data ini dihasilkan oleh sensor yang menerima dan mengukur radiasi termal atau sinar matahari dari pantulan benda di permukaan bumi tanpa memancarkan sinyal sendiri. Data ini direkam oleh satelit optik seperti MODIS, Landsat, dan Sentinel.

b. Data Satelit Aktif

Data satelit aktif merupakan data yang tidak bergantung pada pancaran energi matahari. Data dihasilkan dari sensor yang memancarkan sinyal sendiri, kemudian diukur berdasarkan pantulan dari benda di permukaan bumi. Sensor ini biasanya menggunakan teknologi radar untuk mengumpulkan data tentang struktur, topografi, dan perubahan permukaan bumi. Contoh pada Sentinel-1, data direkam oleh satelit *Synthetic Aperture Radar* (SAR) yang menyediakan data beresolusi tanpa terganggu dengan kondisi cuaca.

Berdasarkan area observasi, data dapat dibagi menjadi tiga, yaitu:

a. Observasi Daratan

Observasi daratan merupakan data yang merekam pada bagian daratan di permukaan bumi. Contoh penggunaan data observasi daratan adalah penggunaan citra Landsat 8 OLI untuk menganalisis tutupan lahan dan penggunaan data MODIS untuk analisis spasial *temporal Land Surface Temperature (LST)*.

b. Observasi Perairan/Lautan

Observasi perairan/lautan merupakan data yang merekam pada bagian perairan dan lautan. Contoh penggunaan data observasi perairan/lautan adalah penggunaan data Landsat-8 untuk menentukan kesesuaian lokasi marikultur ikan kerapu.

c. Observasi Atmosfer

Observasi atmosfer merupakan data yang merekam pada bagian atmosfer. Contoh penggunaan data observasi atmosfer adalah penggunaan data sentinel-4 untuk menganalisis kualitas udara.

Berdasarkan waktunya, data dibagi menjadi menjadi dua, yaitu:

a. Observasi Siang Hari

Data observasi siang hari sebagian besar dilakukan oleh satelit pasif atau satelit opti seperti Landsat dan Sentinel.

b. Observasi Malam Hari

Data observasi malam hari dilakukan oleh satelit aktif untuk mendeteksi Cahaya buatan di permukaan bumi. Data ini biasanya digunakan untuk

memantau urbanisasi dan populasi. Contoh data observasi malam hari adalah *Visible Infrared Imaging Radiometer Suite* (VIIRS) dan *Defense Meteorological Satellite Program- Operational Linescan System* (DMSP-OLS).

Selain kategori diatas, terdapat juga data lainnya seperti data batas administrasi multilevel (FAO GAUL: *Global Administrative Unit Layers*) dan data curah hujan (GPM, TRMM, dan GSMaP) yang dapat dimanfaatkan untuk membantu penelitian.

Pengolahan data di GEE menggunakan bahasa pemrograman Javascript dengan penyimpanan berbasis *cloud computing*. Penggunaan bahasa javascript ini dapat memudahkan pengerjaan karena memiliki sintaks yang sederhana sehingga dapat dengan mudah dipahami. Sementara itu, penyimpanan GEE yang berbasis *cloud computing* turut mendukung dalam melakukan analisis meskipun data yang diproses berukuran besar, tidak diperlukan perangkat keras yang kuat dengan perangkat lunak terbaru. Pemrograman pada GEE juga dilengkapi algoritma untuk melakukan analisis data spasial geografis berukuran besar dan lingkungan pemrograman yang interaktif. Data dalam GEE memiliki rentang waktu yang panjang dan dapat ditampilkan dalam dua dimensi. Hasil pengolahan ini nantinya dapat digunakan untuk penelitian yang berkaitan dengan perubahan lingkungan, kondisi atmosfer, penggunaan lahan, perubahan iklim, urbanisasi, dan lain-lainnya (Zhao et al., 2021).

2.2.3 StandardScaler

StandardScaler merupakan *preprocessing technique* dengan mengubah distribusi fitur dataset menjadi *standard distribution* yang memiliki *mean* (rata-rata)

nol dan *standard deviation* satu (Masacgi & Rohman, 2023). StandardScaler tidak mengubah bentuk distribusi aslinya, hanya menstandarkan skala atribut saja. Standarisasi atribut ini bekerja dengan cara mengurangi setiap nilai individu fitur menggunakan nilai rata-rata dari fitur tersebut. Selanjutnya dibagi dengan standar deviasi dari nilai fitur. Akhirnya nilai setiap fitur akan memiliki *mean* 0 dan *standard deviation* 1. Sehingga dapat mempercepat perhitungan dalam algoritma dan menghasilkan hasil akhir yang lebih baik. Rumus menggunakan metode StandardScaler

$$X_{standardscaler} = \frac{(x - x_{mean})}{standard\ deviation(x)} \quad (2.1)$$

2.2.4 Linear Regression

Linear Regression merupakan metode yang difungsikan untuk mengindikasikan hubungan antar variabel. Selanjutnya hubungan ini dinyatakan dalam bentuk sebuah persamaan. Dalam *linear regression*, terdapat dua tipe variabel, yakni variabel dependen dan variabel independen. Variabel dependen merupakan variabel yang tidak bebas karena nilainya dipengaruhi oleh nilai variabel lain. Variabel dependen biasanya sering disimbolkan dengan y. Sementara itu, variabel independen merupakan variabel yang bebas karena nilainya tidak dipengaruhi oleh nilai variable lain. Variabel independen biasanya disimbolkan dengan x (Latief & Karyanti, 2022). Persamaan dari *linear regression* sederhana sebagai berikut:

$$y = a + bx \quad (2.2)$$

Keterangan

y = variabel dependen atau variabel terikat

x = variabel independent atau variabel bebas

a = konstanta

b = koefisien regresi

Linear Regression juga dapat digunakan pada data yang memiliki satu variabel dependen dan lebih dari satu variabel independen. Model persamaan dari *linear regression* dengan variabel jamak ini sebagai berikut,

$$y = \alpha + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n \quad (2.3)$$

Keterangan

y = nilai variabel dependen saat observasi ke- i

x = nilai variabel independen saat observasi ke- 1

α = konstanta

β = koefisien regresi

Nilai dari konstanta dan koefisien regresi dapat ditentukan dari persamaan:

$$\alpha = \frac{(\sum y_i)(\sum x_i^2) - (\sum x_i)(\sum x_i y_i)}{n \sum x_i^2 - (\sum x_i)^2} \quad (2.4)$$

$$\beta = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2} \quad (2.5)$$

2.2.5 Evaluasi Metrik

Evaluasi metrik merupakan tahapan yang penting saat melakukan analisis regresi maupun *machine learning* lainnya yang menggunakan model prediksi (Plevris et al., 2022). Evaluasi metrik digunakan untuk mengetahui keberhasilan model prediksi yang dibangun dengan menghitung kedekatan hasil prediksi dan data sebenarnya. (Yusuf & Huda, 2023). Evaluasi matrik yang digunakan untuk

mengukur model antara lain *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE) dan *Root Mean Squared Error* (RMSE).

2.2.5.1 Mean Absolute Error (MAE)

Mean Absolute Error (MAE) merupakan *evaluation matrix* yang digunakan untuk mengukur kedekatan hasil prediksi dengan nilai aktual. MAE memberikan gambaran besar kesalahan prediksi suatu data tanpa memperhatikan arah kesalahannya. Arah kesalahan yang dimaksud adalah nilai positif dan negatif (Amansyah et al., 2024). Evaluasi ini menghitung dengan menemukan selisih absolut dari nilai hasil prediksi dengan nilai yang sebenarnya lalu mengambil rata-rata dari hasil selisih tersebut. MAE baik digunakan untuk pengukuran prediksi *error* pada analisis *time series* (Pebralia, 2022). Persamaan 2.6 berikut merupakan cara untuk menghitung MAE:

$$MAE = \sum \frac{|y' - y|}{n} \quad (2.6)$$

2.2.5.2 Mean Squared Error (MSE)

Mean Squared Error (MSE) merupakan *evaluation matrix* yang digunakan untuk mengukur kinerja model dengan menghitung kuadrat dari selisih antara hasil prediksi dengan nilai data aktual. Jika hasil MSE tinggi berarti hasil prediksi model sering berbeda jauh dari nilai data aktual (Arisandi et al., 2023). Sehingga model tersebut tidak konsisten dalam melakukan prediksi. Persamaan 2.7 berikut merupakan cara untuk menghitung MSE:

$$MSE = \sum \frac{(y' - y)^2}{n} \quad (2.7)$$

2.2.5.3 Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) merupakan *evaluation matrix* yang berasal dari akar kuadrat dari MSE. RMSE menghasilkan pengukuran kesalahan prediksi di unit yang sama dengan variabel target. RMSE akan selalu bernilai positif. Semakin rendah atau mendekati nol suatu hasil RMSE menunjukkan hasil prediksi yang dihasilkan semakin sesuai dengan data aktual. Sebaliknya jika hasil RMSE *relative* besar maka hasil prediksi yang dihasilkan semakin jauh dari nilai sebenarnya (Vermaysha & Nurmalitasari, 2023). Persamaan 3.3 berikut merupakan cara untuk menghitung RMSE:

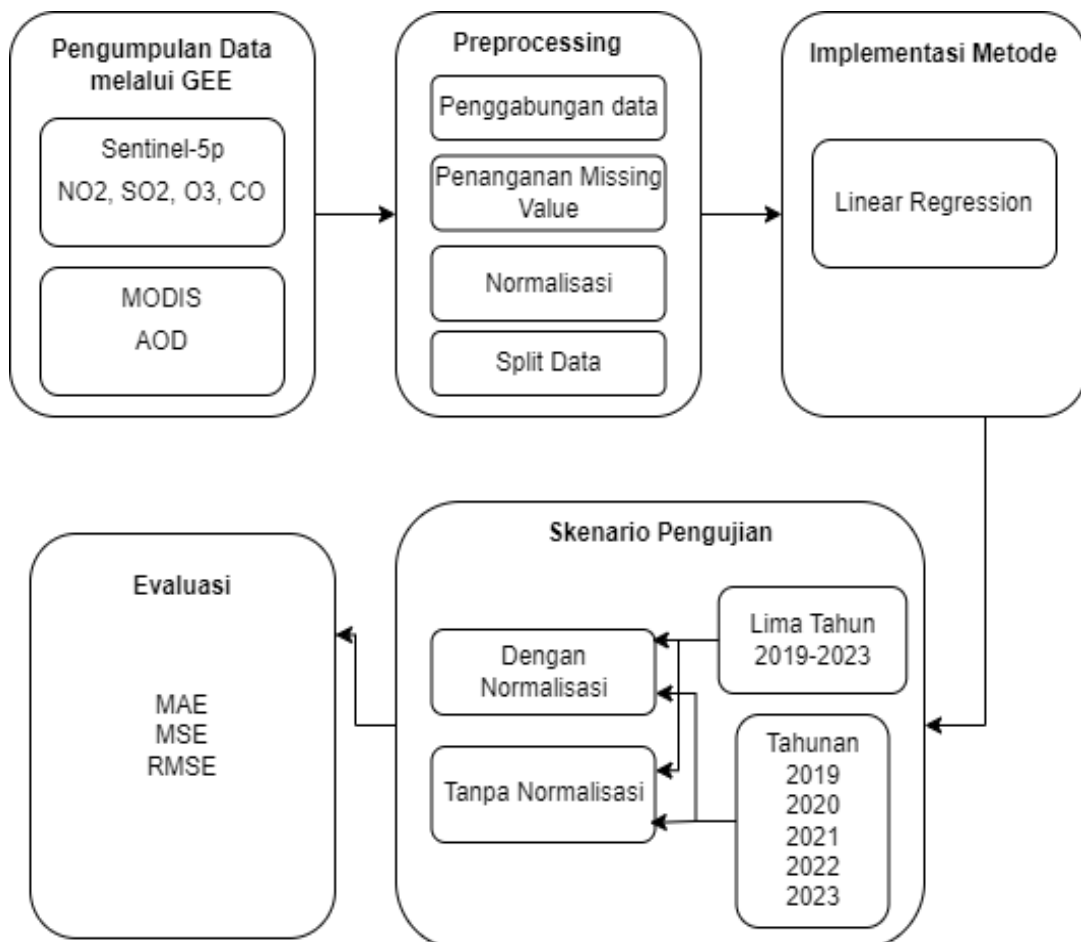
$$RMSE = \sqrt{\sum \frac{(y' - y)^2}{n}} \quad (2.7)$$

BAB III

DESAIN PENELITIAN

3.1 Desain Alur Penelitian

Desain alur penelitian merupakan rangkaian proses penelitian dari awal hingga akhir. Desain dibuat agar penelitian menjadi lebih terarah dan sistematis. Dalam setiap subbab ini akan menguraikan tentang *step* yang akan dijalankan dalam proses prediksi kualitas udara menggunakan *linear regression* seperti pada gambar 3.1

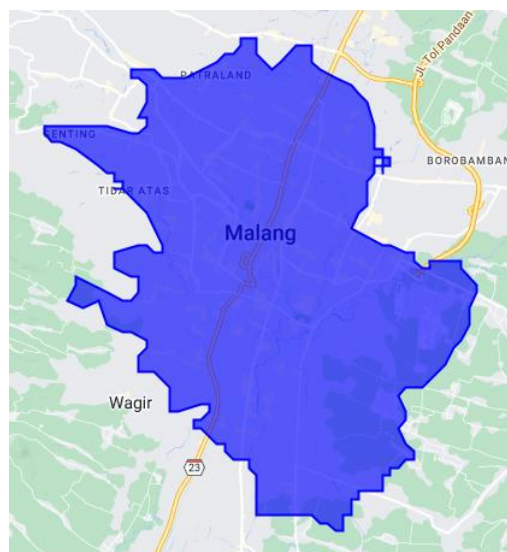


Gambar 3. 1 Alur Penelitian

3.2 Pengumpulan Data Melalui GEE

Penelitian ini menggunakan data yang dikumpulkan dari Sentinel-5 dan MODIS melalui GEE. Pengolahan data menggunakan GEE yang berbasis *cloud computing*, membuat pengumpulan data menjadi lebih cepat dan mutakhir. Hal ini dikarenakan tanpa mengunduh berbagai macam citra sudah dapat dilakukan pemrosesan hanya dengan adanya akses internet (Fikri et al., 2022). Selain itu, GEE juga menyediakan berbagai macam jenis data dalam rentang waktu yang lama hingga *real-time* data.

Data yang digunakan dipilih dengan menggunakan beberapa filter. Filter *date* untuk membatasi rentang waktu data yang akan digunakan. Pada penelitian ini waktu yang digunakan dimulai dari 1 Januari 2019 hingga 31 Desember 2023. Filter *bound* digunakan untuk membuat batas daerah yang diteliti. Dalam penelitian ini daerah yang diteliti adalah kota Malang. Filter *reduce* digunakan untuk menentukan bagaimana data dikumpulkan. Dalam penelitian ini, data dikumpulkan dengan rata-rata per hari.



Gambar 3. 2 Batas Wilayah Kota Malang

Penentuan parameter yang digunakan pada penelitian ini mempertimbangkan ketentuan komponen penentu kualitas udara dari beberapa pihak. Yang pertama, berdasarkan Peraturan Menteri Lingkungan Hidup Dan Kehutanan Republik Indonesia Nomor P.14/MENLHK/SETJEN/KUM.1/7/2020 Tentang Indeks Standar Pencemar Udara, (2020) menetapkan parameter yang digunakan untuk mengetahui indeks kualitas udara adalah partikulat (PM_{10}), partikulat ($PM_{2.5}$), karbon monoksida (CO), nitrogen dioksida (NO_2), sulfur dioksida (SO_2), dan ozon (O_3). Yang kedua, menurut Peraturan Badan Meteorologi, Klimatologi, dan Geofisika Republik Indonesia Nomor 2 Tahun 2020 Tentang Penyediaan dan Penyebaran Informasi Kualitas Udara, (2020) parameter kualitas udara yang digunakan adalah partikulat (PM_{10} , $PM_{2.5}$), sulfur dioksida (SO_2), nitrogen oksida (NO_2), ozon (O_3), karbon monoksida (CO), karbon dioksida (CO_2), dan metan (CH_4). Yang ketiga, AQAir menetapkan enam parameter yang digunakan untuk mengukur kualitas udara enam parameter tersebut meliputi, $PM_{2.5}$, PM_{10} , karbon monoksida (CO), sulfur dioksida (SO_2), nitrogen dioksida (NO_2), dan ozon (O_3). AQAir sendiri merupakan perusahaan teknologi yang bergerak dalam penyampaian informasi terkait kualitas udara dengan memberdayakan pemerintah, organisasi, maupun individu. Di Indonesia, KLHK dan BMKG menjadi kontributor terbanyak untuk AQAir Indonesia. Oleh karena itu, dipilihlah parameter yang ada dalam ketentuan komponen penentu kualitas udara seperti yang telah dijabarkan sebelumnya dan datanya tersedia pada GEE sehingga parameter yang digunakan adalah karbon monoksida (CO), sulfur

dioksida (SO₂), nitrogen dioksida (NO₂), ozon (O₃), dan PM_{2.5} yang menggunakan nilai AOD.

3.3 Preprocessing

Preprocessing merupakan tahap awal untuk mempersiapkan data agar data siap untuk dilakukan pengujian. Pada tahap ini dibagi menjadi 4 tahapan, yaitu penggabungan data, penanganan missing value, normalisasi, dan split data.

3.3.1 Penggabungan Data

Variabel yang didapatkan dari GEE memiliki beberapa hari tanpa hasil pengukuran. Untuk mengatasi hal tersebut, data di digabungkan berdasarkan kesamaan tanggal pengukuran parameter AOD. Hal ini dikarenakan AOD akan menjadi variabel dependennya. Daftar parameter beserta keterangan yang digunakan dalam penelitian ini tercantum pada tabel 3.1.

Tabel 3. 1 Keterangan Parameter

No.	Parameter	Keterangan
1.	Tanggal	Tanggal pengambilan data
2.	CO	Carbon Monoksida
3.	NO ₂	Nitrogen Dioksida
4.	SO ₂	Sulfur Dioksida
5.	O ₃	Ozon
6.	AOD	<i>Aerosol Optical Depth</i>

3.3.2 Penanganan *Missing Value*

Setelah penggabungan data, terdapat data yang kosong. Kosongnya data disebabkan pada tanggal tersebut satelit tidak melakukan pengukuran. Adanya data kosong dalam dataset dapat dapat menurunkan kualitas data sehingga ketika dilakukan pemrosesan, dataset tersebut dapat mengacaukan hasil prediksi (Khan &

Hoque, 2020). Untuk mengatasi permasalahan ini dapat dilakukan dengan menggunakan teknik imputasi. *Mean* atau rata-rata data menjadi salah satu teknik imputasi yang banyak digunakan. Penggunaan *mean* pada dataset *time series* lebih baik daripada teknik komputasi konvensional lainnya (Prasetya et al., 2023).

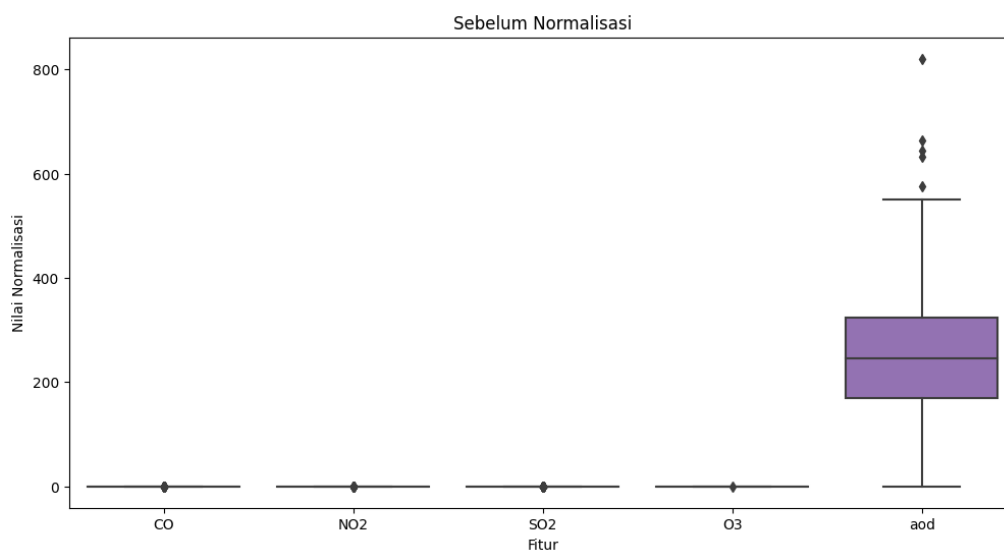
Imputasi pada dataset dilakukan pada tiap parameternya, misalkan terdapat data kosong pada parameter CO, maka dalam tahap ini, nilai kosong tersebut akan diisi dengan nilai rata-rata dari keseluruhan nilai CO. Hal ini juga berlaku untuk parameter NO₂, SO₂, dan O₃. Perubahan nilai kosong pada data lima tahun disetiap parameter termuat dalam tabel 3.2 yang ditandai warna abu-abu untuk CO, merah untuk NO₂, kuning tua untuk SO₂, dan biru tua untuk O₃. Sebagai contoh pada baris pertama kolom CO semula tidak memiliki nilai berubah menjadi 0,028867 mol/m² yang merupakan nilai rata-rata dari parameter CO. Selanjutnya, parameter SO₂ dan O₃ pada baris yang sama berubah menjadi 0,000488 mol/m² dan 0,114651 mol/m². Sama halnya dengan baris pertama, pada baris terakhir pada parameter NO₂ yang berubah menjadi 6,81E-05 mol/m².

Tabel 3. 2 Data Setelah Penanganan *Missing Value*

Tanggal	CO	NO2	SO2	O3	AOD
2019-01-12	0,028867	4,74E-05	0,000488	0,114651	292
2019-01-15	0,021633	0,000025	0,000488	0,117673	265,077
2019-01-20	0,026753	4,76E-05	0,000488	0,115193	209,406
...
2023-12-22	0,026381	5,05E-05	0,000488	0,113733	89,437
2023-12-23	0,026771	6,81E-05	0,000488	0,11294	174,716
2023-12-27	0,027388	0,000053	0,000488	0,113915	322

3.3.3 Normalisasi

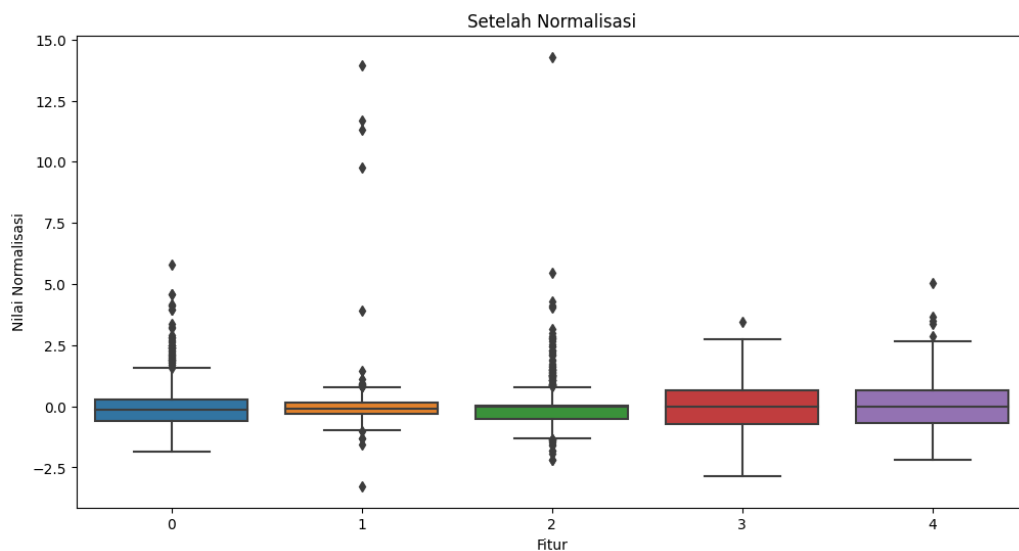
Data yang telah terisi semua dalam setiap selnya ternyata memiliki rentang yang jauh antar parameternya. Sehingga diperlu dilakukan normalisasi agar tidak ada dominasi antar variabel. Normalisasi merupakan proses mengubah nilai aktual menjadi nilai yang memiliki rentang baru (Ginting et al., 2021). Pada penelitian ini menggunakan StandardScaler untuk menormalisasi data. Penggunaan StandarScaler untuk *linear regression* dikarenakan *linear regression* memodelkan keterkaitan antara variabel independen dengan variabel dependen. Jika terdapat variabel yang memiliki nilai dengan skala yang lebih besar daripada variabel lainnya, maka variabel tersebut akan mendominasi.



Gambar 3. 3 Boxplot Data Sebelum Normalisasi

Gambar 3.3 merupakan data yang belum dilakukan normalisasi, komponen *boxplot* pada parameter AOD terlihat dengan sangat jelas dibandingkan dengan variabel lainnya. Hal ini dikarenakan parameter tersebut memiliki rentang nilai yang lebih besar dibandingkan dengan nilai parameter lainnya yang hanya berada

di rentang -0,001 hingga 1. Setelah dilakukan normalisasi seperti pada gambar 3.4, nilai antar parameter akan memiliki rentang yang baru. Nilai setiap parameter akan memiliki nilai *mean* nol dengan *deviation standard* satu.



Gambar 3. 4 Boxplot Data Sesudah Normalisasi

3.3.4 Split Data

Split data sangat diperlukan dalam penelitian ini. Komposisi *split data* dapat mempengaruhi hasil dari kinerja model dan dapat menghindari *overfitting* maupun *underfitting* (Oktafiani et al., 2023). Saat melakukan *split data*, data dipisahkan menjadi dua kategori, yaitu data *training* dan data *testing*. Model dibangun dengan menggunakan data *training* dan diuji menggunakan data *testing*. Banyak komposisi yang dapat digunakan untuk membagi data *training* dan data *testing* antara lain, 50%:50%, 60%:40%, 70%:30%, 73%:27%, 75%:25%, 80%:20%, 83%:17%, 85%:15%, dan 90%:10% (Azmi et al., 2023). Akan tetapi, sebagian besar penelitian lebih banyak menggunakan komposisi 75%:25%, 80%:20%, dan 90%:10%.

Langkah awal dari tahap ini yaitu dilakukan pemisahan antara variabel dependen dan independen. Parameter AOD digunakan sebagai variabel dependen, sedangkan variabel independennya menggunakan parameter NO₂, SO₂, O₃, dan CO. Selanjutnya, data akan dibagi menjadi data *training* dan data *testing*.

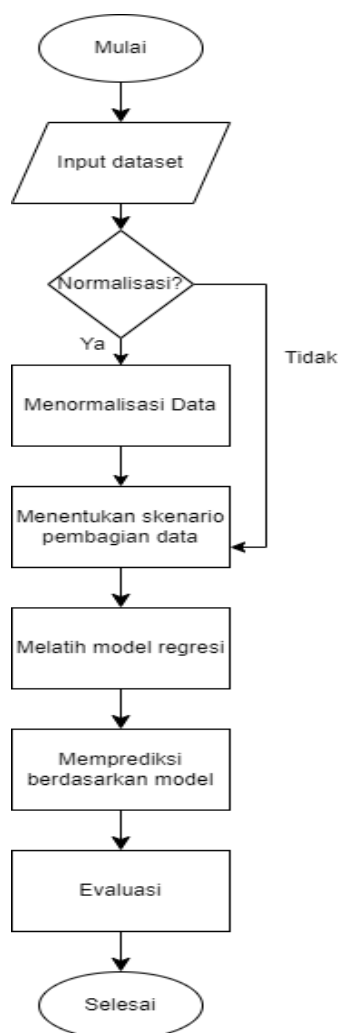
Split data pada penelitian ini dilakukan dengan menggunakan tiga model seperti yang tercantum pada tabel 3.3. Model pertama data dipisah menjadi 75% data *training* dan 25% data *testing*. Model kedua data dipisah menjadi 80% data *training* dan 20% data *testing*. Model ke tiga data dipisah menjadi 90% data *training* dan 10% data *testing*.

Tabel 3. 3 Model *Split* Data

Model	Komposisi Pembagian Data	
	<i>Training</i>	<i>Testing</i>
Model 1	75%	25%
Model 2	80%	20%
Model 3	90%	10%

3.4 Implementasi Metode

Implementasi metode *linear regression* ini digunakan untuk melakukan prediksi. Karena variabel independen yang akan diproses lebih dari satu, maka dipilihlah metode *Linear Regression* dengan variabel jamak. Proses implementasi metode tersaji dalam bentuk *flowchart* pada gambar 3.5

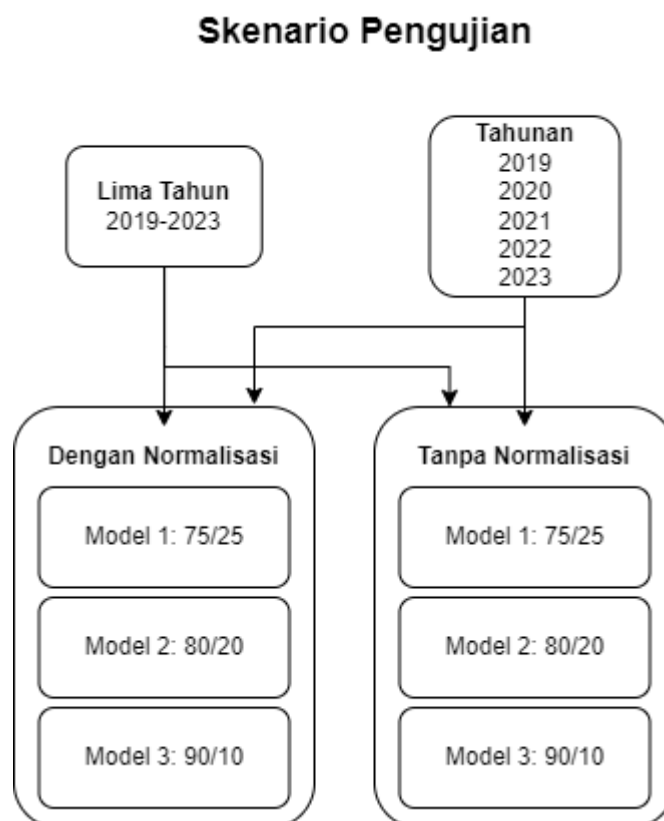


Gambar 3. 5 Flowchart Metode *Linear Regression*

Implementasi metode *Linear Regression* dimulai dengan menginputkan dataset. Data yang diaplikasikan adalah data yang telah ditangani *missing value*-nya atau data yang tidak memiliki *cell* kosong lagi. Selanjutnya, menentukan skenario pengujian dengan memilih antara melakukan normalisasi atau tidak terhadap data. Jika skenario yang dipilih menggunakan normalisasi, maka data akan dinormalisasi dahulu. Jika skenario yang dipilih tanpa normalisasi maka dapat lanjut pada tahap berikutnya. Tahap berikutnya adalah penentuan model pembagian komposisi data. Pemilihan model dilakukan sesuai dengan *split data* pada tabel 3.3

di *subbab* sebelumnya. Setelah dilakukan pembagian model, selanjutnya model *linear regression* akan dilatih menggunakan data *training* sesuai dengan model pembagian data. Dari pelatihan model ini didapatkan konstanta dan koefisien regresi. Konstanta dan koefisien regresi ini selanjutnya digunakan untuk memprediksi pada tahap berikutnya. Kemudian dilakukan evaluasi pada hasil prediksi dengan data sebenarnya. Evaluasi menggunakan *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE). Dengan demikian, implementasi Metode *Linear Regression* selesai.

3.5 Skenario Pengujian



Gambar 3. 6 Skenario Pengujian

Pada penelitian ini dilakukan dengan menerapkan dua skenario pengujian. Skenario pertama menggunakan data selama lima tahun (2019-2023) dan skenario kedua menggunakan data pertahun mulai dari 2019, 2020, 2021, 2022, dan 2023. Pada masing-masing skenario akan dilakukan pengujian dengan data yang dinormalisasi dan data tanpa normalisasi. Disetiap pengujian terdapat tiga model rasio yang akan diuji berdasarkan *split data* pada tabel 3.3. Skenario yang digunakan dalam penelitian ini dapat dilihat pada gambar 3.6. Dengan adanya perbedaan skenario ini akan ditemukan perbedaan hasil prediksi, sehingga dapat dipilih hasil terbaik.

3.6 Evaluasi

Setelah didapatkan hasil prediksi berdasarkan skenario pengujian, selanjutnya dilakukan evaluasi pada model prediksi. Evaluasi diperlukan untuk mengetahui kedekatan antara hasil prediksi dan nilai sebenarnya. Pada penelitian ini, menggunakan evaluasi metrik yaitu *Mean Absolut Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE).

BAB IV

UJI COBA DAN PEMBAHASAN

4.1 Pengumpulan Data

Pengumpulan data dilakukan melalui *Google Earth Engine* (GEE) menggunakan dua satelit yaitu Sentinel-5 dan MODIS. Data dikumpulkan berdasarkan rata-rata harian polutan dari tanggal 1 Januari 2019 hingga 31 Desember 2023 dengan batas daerah penelitian dalam Kota Malang.

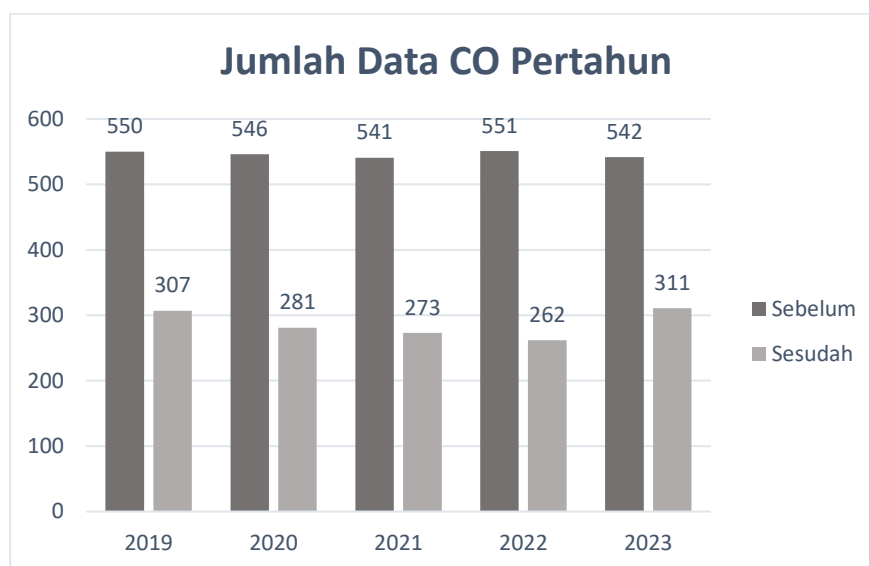
Selanjutnya, data dikumpulkan berdasarkan parameter secara terpisah. Hal ini dikarenakan tiap parameter memiliki jumlah data pengukuran yang berbeda dengan parameter lainnya, sehingga tidak dapat langsung diimport dalam satu dataset. Perbedaan jumlah data pengukuran ini disebabkan oleh adanya hari tanpa pengukuran dan duplikasi tanggal pengukuran. Duplikasi tanggal pengukuran tersebut banyak yang bernilai kosong. Seperti pada parameter CO pada tabel 4.1, pada baris yang berwarna hijau. Data yang kosong tersebut selanjutnya dihapus agar tidak mempengaruhi proses selanjutnya. Penghapusan data ini memberikan selisih yang banyak antara data awal yang di *import* dari GEE dengan data setelah penghapusan.

Tabel 4. 1 Data Awal Parameter CO

Tanggal	Rata-rata harian (mol/m ²)
2019-01-01	0,0282883
2019-01-02	0,0258516
2019-01-03	
2019-01-03	
2019-01-04	0,0252889
2019-01-05	0,0280081

2019-01-06	0,0302191
2019-01-07	0,0311293
2019-01-08	
2019-01-08	0,0262540

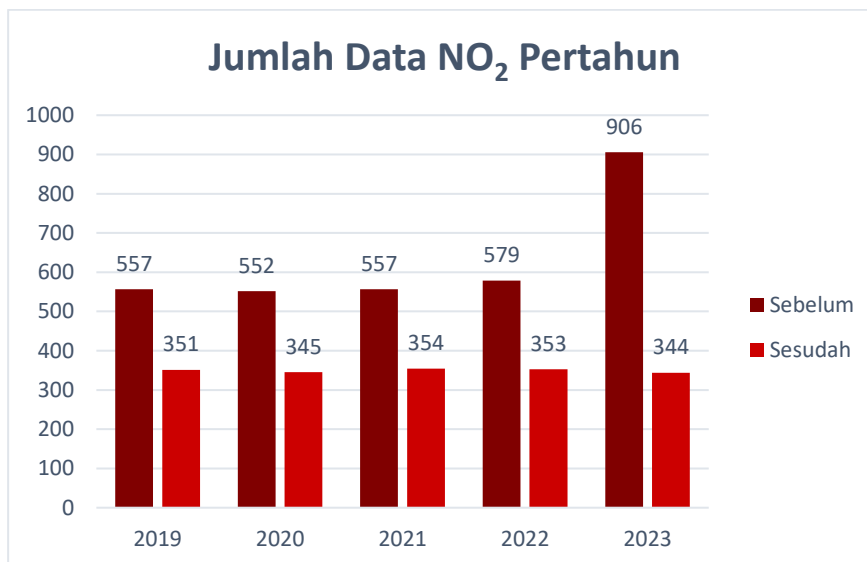
Pada pengumpulan data parameter CO, data awal memiliki jumlah data pertahun yang *relative* sama, yaitu pada angka 500-an. Setelah dilakukan penghapusan data kosong, didapatkan hasil yang bervariasi. Data dengan jumlah paling sedikit ada pada tahun 2022 dengan 262 baris data dan terbanyak ada pada tahun 2023 sebanyak 311 baris data. Perbandingan antara jumlah data awal dengan data setelah penghapusan tersaji pada gambar 4.1.



Gambar 4. 1 Perbandingan Jumlah Baris Data CO

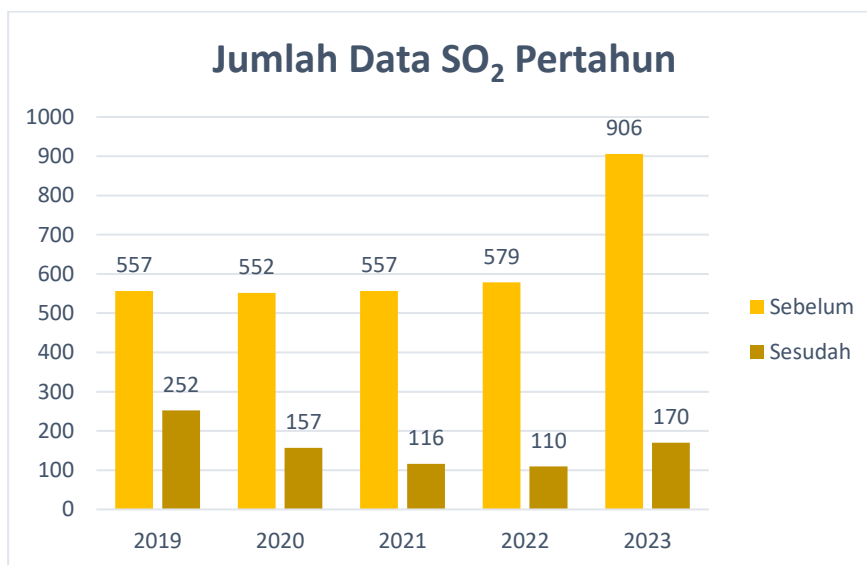
Pada pengumpulan data parameter NO₂, data awal memiliki jumlah data pertahun yang *relative* sama pada tahun 2019-2022, yaitu pada angka 500-an. Akan tetapi pada tahun 2023 jumlah data melonjak menjadi 906 baris data. Setelah dilakukan penghapusan data kosong, didapatkan hasil yang *relative* sama tiap tahunnya. Banyak data berada pada nilai mulai 344, yang berarti pada tiap tahun

satelit banyak melakukan pengukuran. Perbandingan antara jumlah data awal dengan data setelah penghapusan tersaji pada gambar 4.2.



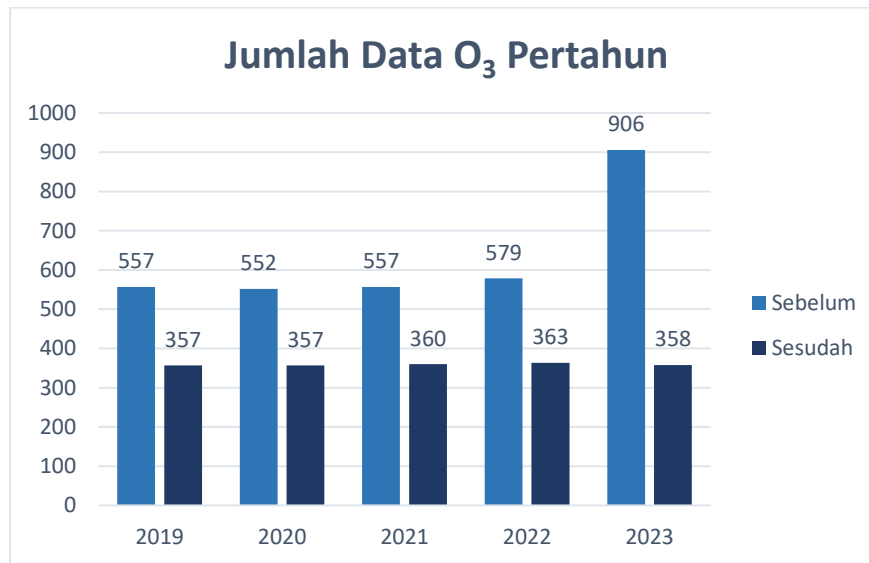
Gambar 4. 2 Perbandingan Jumlah Baris Data NO₂

Pada pengumpulan data parameter SO₂, data awal memiliki jumlah data pertahun yang *relative* sama pada tahun 2019-2022, yaitu pada angka 500-an. Akan tetapi pada tahun 2023 jumlah data melonjak menjadi 906 baris data. Setelah dilakukan penghapusan data kosong, didapatkan hasil yang bervariasi. Data pada parameter SO₂ ini memiliki selisih yang lebih dari setengah data awal. Data dengan jumlah paling sedikit terdapat pada tahun 2022 dengan 110 baris data dan terbanyak ada pada tahun 2019 sebanyak 252 baris data. Perbandingan antara jumlah data awal dengan data setelah penghapusan tersaji pada gambar 4.3.



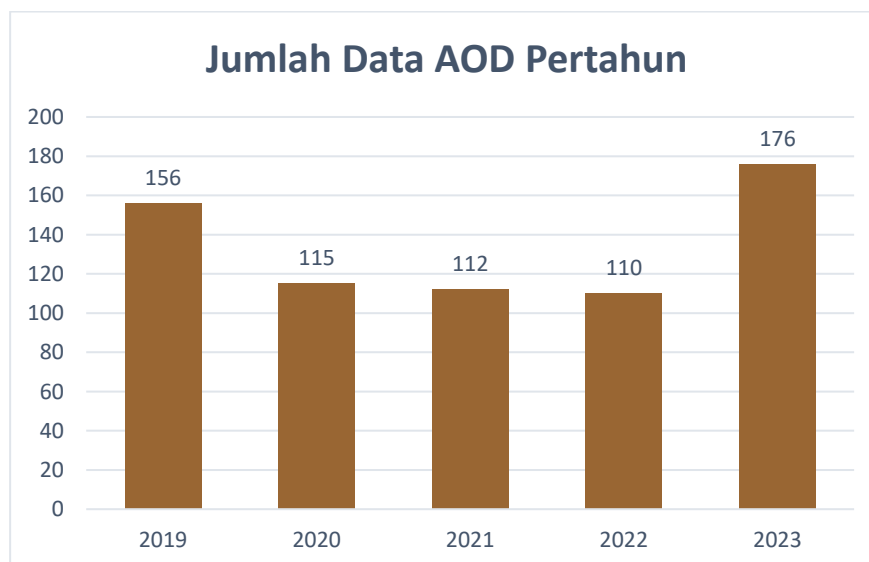
Gambar 4. 3 Perbandingan Jumlah Baris Data SO₂

Pada pengumpulan data parameter O₃, data awal memiliki jumlah data pertahun yang *relative* sama pada tahun 2019-2022, yaitu pada angka 500-an. Akan tetapi pada tahun 2023 jumlah data melonjak menjadi 906 baris data. Setelah dilakukan penghapusan data kosong, didapatkan hasil yang *relative* sama tiap tahunnya. Data dengan jumlah paling sedikit terdapat pada tahun 2019 dan 2020 dengan 357 baris data dan terbanyak ada pada tahun 2022 sebanyak 363 baris data. Perbandingan antara jumlah data awal dengan data setelah penghapusan tersaji pada gambar 4.4.



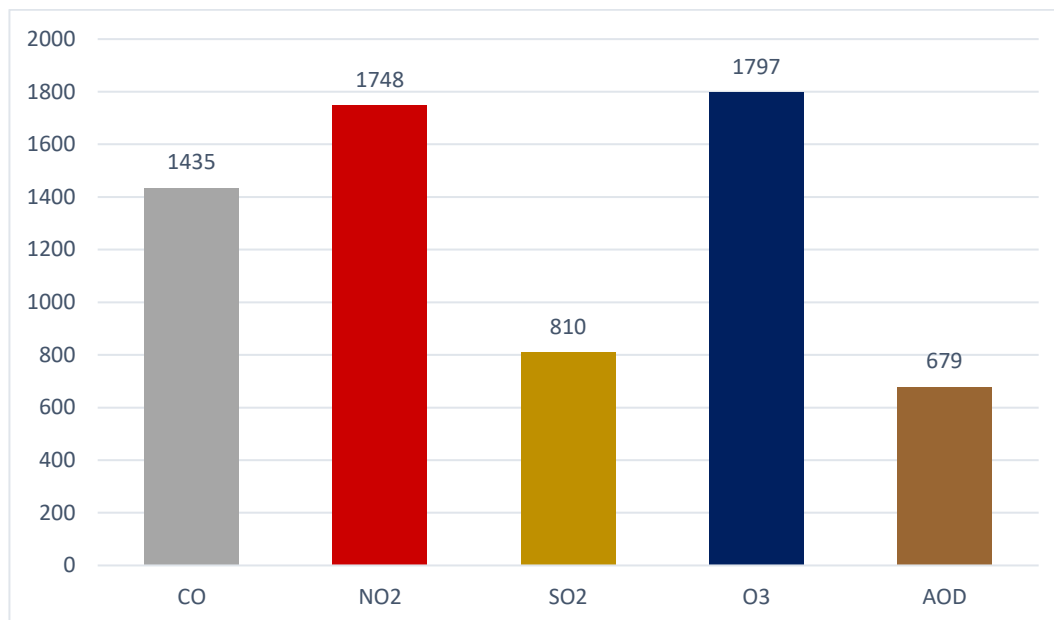
Gambar 4. 4 Perbandingan Jumlah Baris Data O₃

Pada pengumpulan data AOD tidak ditemukan data yang kosong dan terduplikasi, sehingga tidak diperlukan penghapusan. Parameter AOD memiliki jumlah data terbanyak pada tahun 2019 dan jumlah data paling sedikit tahun 2022. Jumlah data AOD pertahun dimuat pada gambar 4.5.



Gambar 4. 5 Perbandingan Jumlah Baris Data AOD

Selanjutnya, jumlah total selama lima tahun tiap parameter dimuat pada gambar 4.6 berikut.



Gambar 4. 6 Perbandingan Jumlah Baris Data Tiap Parameter Setelah Penghapusan Data Kosong

4.2 Preprocessing

4.2.1 Penggabungan Data

Penggabungan data dilakukan di excel menggunakan formula VLOOKUP. VLOOKUP akan mencari data di sheet yang berbeda berdasarkan kesamaan data yang diketahui secara vertikal sehingga data yang pertama kali ditemukan akan menjadi hasil dari formula ini (Nurramadhani et al., 2024). Dalam penelitian ini, data digabungkan berdasarkan kesamaan tanggal pengambilan data parameter AOD. Hasil dari penggabungan data dapat dimuat pada tabel 4.1.

Tabel 4. 2 Penggabungan Data Awal

N o.	Tanggal	CO	NO2	SO2	O3	AOD
1.	2019-01-12	#N/A	4.746760079983625E-5	#N/A	#N/A	292,000
...
157.	2020-01-14	0,0281370	6.1186810483075E-5	1.1946358998095884E-4	0,1156988	413,908
...
272.	2021-01-24	#N/A	3.317799137189506E-5	#N/A	0,1108321	433,000
...
393.	2022-01-03	0,0254302	5.263555455821654E-5	4.6217269862829563E-4	0,1155982	574,946
...
504.	2023-01-04	#N/A	5.362720621384728E-5	#N/A	0,1195771	138,951
....
520.	2023-03-15	#N/A	#N/A	#N/A	#N/A	359,758
521.	2023-03-17	#N/A	#N/A	#N/A	#N/A	455,131
...
679.	2023-12-27	0,0273881	5.308374756552486E-5	#N/A	0,1139155	322,000

Setelah data digabungkan terdapat hasil #N/A yang berarti baris data dengan tanggal yang dicari tidak ditemukan atau data tidak tersedia pada tanggal tersebut. Jika suatu baris data memiliki hasil #N/A pada semua kolom parameter independen (CO, NO2, SO2, O3) seperti pada baris nomor 520 dan 521, maka baris data tersebut akan dihapus. Hal ini dikarenakan penelitian ini berdasarkan polutan dan untuk memaksimalkan hasil prediksi.

Selanjutnya, dilakukan penyeragaman variasi penulisan angka untuk menjaga konsistensi pada data. Data yang konsisten dapat terhindar dari disfungsi yang tidak diinginkan pada saat pemrosesan (Ayuningtyas & Yustanti, 2024).

Penyeragaman format data menggunakan bentuk bilangan desimal dikarenakan tiga dari 5 kolom parameter menggunakan bentuk bilangan desimal. Sehingga kolom NO₂ dan SO₂ diubah menjadi bilangan desimal dengan tujuh angka di belakang nol seperti kolom CO dan O₃.

Kemudian didapatkanlah hasil akhir data yang akan digunakan untuk penelitian ini dengan total 669 baris data. Hasil akhir dari proses penggabungan data termuat pada tabel 4.3.

Tabel 4. 3 Penggabungan Data Akhir

No.	Tanggal	CO	NO ₂	SO ₂	O ₃	AOD
1.	2019-01-12	nan	0,0000474	nan	nan	292,000
...
157.	2020-01-14	0,0281370	0,0000611	0,0001194	0,1156988	413,908
...
272.	2021-01-24	nan	0,0000331	nan	0,1108321	433,000
...
390.	2022-01-03	0,0254302	0,0000526	0,0004621	0,1155982	574,946
...
500.	2023-01-04	nan	0,0000536	nan	0,1195771	138,951
...
669.	2023-12-27	0,0273881	0,0000530	nan	0,1139155	322,000

4.2.2 Penanganan *Missing Value*

Pada hasil akhir dari penggabungan data pada tabel 4.3 terdapat nilai kosong pada beberapa parameter. Nilai kosong sebuah parameter diisi menggunakan nilai rata-rata dari parameter tersebut. Hal ini dikarenakan penggunaan mean pada dataset *time series* lebih baik daripada teknik komputasi

konvensional lainnya (Prasetya et al., 2023). Seperti pada data lima tahun, nilai rata-rata parameter CO sebesar $0,028867 \text{ mol/m}^2$, NO₂ sebesar $6,81\text{E-}05 \text{ mol/m}^2$, SO₂ sebesar $0,000488 \text{ mol/m}^2$, dan O₃ sebesar $0,114651 \text{ mol/m}^2$. Hasil dari penanganan *missing value* dimuat dalam tabel 4.4.

Tabel 4. 4 Data Setelah Penanganan *Missing Value*

Tanggal	CO	NO2	SO2	O3	AOD
2019-01-12	0,028867	4,74E-05	0,000488	0,114651	292
...
2020-01-14	0,028137	6,11E-05	0,000119	0,115699	413,908
...
2021-01-24	0,028867	3,31E-05	0,000488	0,110832	433
...
2022-01-03	0,02543	5,26E-05	0,000462	0,115598	574,946
...
2023-01-04	0,028867	5,36E-05	0,000488	0,119577	138,951
...
2023-12-23	0,026771	6,81E-05	0,000488	0,11294	174,716
2023-12-27	0,027388	0,000053	0,000488	0,113915	322

4.2.3 Normalisasi

Normalisasi dilakukan pada semua parameter kecuali kolom tanggal menggunakan `StandardScaler` dari modul `sklearn.preprocessing`. Hasil normalisasi data dimuat dalam tabel 4.4.

Tabel 4. 5 Data Setelah Proses Normalisasi

Tanggal	CO	NO2	SO2	O3	AOD
2019-01-12	0	-0,46207	-1,1E-16	-4,4E-15	0,390765
...
2020-01-14	-0,14117	-0,15666	-0,72963	0,329521	1,459153
...
2021-01-24	0	-0,78085	-1,1E-16	-1,20072	1,626473
...
2022-01-03	-0,66475	-0,34614	-0,05178	0,297891	2,870472
...
2023-01-04	0	-0,32385	-1,1E-16	1,548978	-0,95054

..
2023-12-27	-0,28603	-0,33723	-1,1E-16	-0,23122	0,653681

4.2.4 Split Data

Penelitian ini menggunakan 3 model *split data* dengan rasio yang berbeda-beda. Pada skenario pengujian data lima tahun, model 1 membagi dengan 501 data menjadi *training data* dan 168 menjadi *testing data*. Model 2 membagi 535 menjadi *training data* dan 134 menjadi *testing data*. Model 3 membagi 541 *training data* dan 67 *testing data*. Perbandingan rasio *split data* termuat pada tabel 4.6.

Tabel 4. 6 Model *Split Data* Lima Tahun

Model	Komposisi Pembagian Data	Jumlah Data <i>Training</i>	Jumlah Data <i>Testing</i>	Jumlah Data Total
Model 1	75/25	501	168	669
Model 2	80/20	535	134	669
Model 3	90/10	541	67	669

Pada skenario data tahunan, juga terdapat tiga komposisi pembagian data menjadi tiga model. Data total antar tahun memiliki jumlah yang berbeda, sehingga setiap tahun juga memiliki jumlah data *training* dan data *testing* yang berbeda pula. Jumlah pembagian data *training* dan *testing* tiap tahun ditunjukkan tabel 4.7.

Tabel 4. 7 Model *Split Data* Pertahun

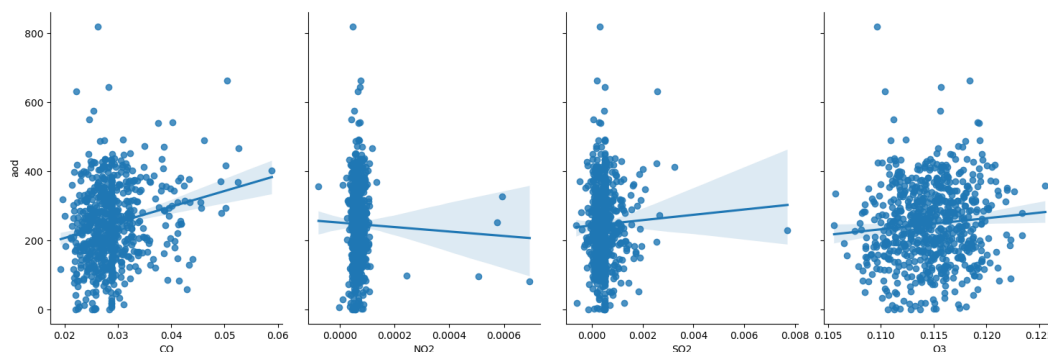
Tahun	Model	Komposisi Pembagian Data	Jumlah Data <i>Training</i>	Jumlah Data <i>Testing</i>	Jumlah Data Total
2019	Model 1	75/25	117	39	156
	Model 2	80/20	124	32	
	Model 3	90/10	140	16	
2020	Model 1	75/25	86	29	115
	Model 2	80/20	92	23	
	Model 3	90/10	103	12	
2021	Model 1	75/25	87	30	117

	Model 2	80/20	93	24	
	Model 3	90/10	105	12	
2022	Model 1	75/25	83	28	111
	Model 2	80/20	88	23	
	Model 3	90/10	99	12	
2023	Model 1	75/25	127	43	170
	Model 2	80/20	136	34	
	Model 3	90/10	153	17	

4.3 Implementasi Metode

Implementasi metode pada tiap prosesnya dilakukan dengan mengacu pada subbab 3.3 pada bab sebelumnya. Metode *linear regression* yang digunakan adalah *linear regression* dengan variabel jamak. Pembuatan model prediksi dengan *linear regression* menggunakan bahasa python dan menggunakan kelas `LinearRegression` dari modul `sklearn.linear_model`. Model tersebut kemudian dilatih untuk menghitung konstanta dan koefisien regresi. Konstanta dan koefisien regresi ini selanjutnya digunakan untuk melakukan prediksi pada data *testing*.

Sebelum proses implementasi, dilakukan uji korelasi untuk mengetahui keterikatan antar variabel. Hubungan antar tiap variabel independen dan variabel dependen terdapat pada gambar 4.7 sedangkan koefisien korelasi tiap variabel terdapat pada tabel 4.8. Koefisien korelasi tidak menggambarkan hubungan sebab akibat melainkan hubungan *linear* antar variabel (Wibowo & Kurniawan, 2020). Nilai koefisien korelasi berada diantara -1 hingga 1. Semakin mendekati 1 baik negatif maupun positif menunjukkan semakin kuat hubungannya dan saat bernilai 0 berarti tidak memiliki hubungan. Berdasarkan tabel tersebut, parameter yang paling berpengaruh terhadap AOD yaitu CO dengan koefisien korelasi sebesar 0,205639.



Gambar 4. 7 Hubungan Antar Variabel Independen Dengan Variabel Dependen

Tabel 4. 8 Hubungan Antar Variabel

Parameter	CO	NO2	SO2	O3	AOD
CO	1.000000	0.083970	-0.103430	0.350327	0.205639
NO2	0.083970	1.000000	0.017986	-0.036954	-0.025307
SO2	-0.103430	0.017986	1.000000	-0.102465	0.034117
O3	0.350327	-0.036954	-0.102465	1.000000	0.089501
AOD	0.205639	-0.025307	0.034117	0.089501	1.000000

4.4 Skenario Pengujian

4.4.1 Data Lima Tahun (2019-2023)

Pada skenario ini, dilakukan pengujian data lima tahun, mulai dari tahun 2019 hingga tahun 2023. Data tersebut akan menggunakan 2 perlakuan pada data, yaitu dengan normalisasi dan tanpa normalisasi. Pada tiap-tiap perlakuan terdapat 3 model yang akan diuji. Sehingga, ada 6 model yang di uji dengan data lima tahun ini.

4.4.1.1 Dengan Normalisasi

Pada pengujian ini ditambahkan normalisasi data menggunakan StandardScaler pada proses *preprocessing*. Terdapat tiga model yang akan diujikan. Perbedaan antar tiga model tersebut didasarkan pada perbandingan rasio

split data. Setelah dilakukan implementasi metode *linear regression* didapatkanlah hasil prediksi. Kemudian hasil prediksi tersebut dikembalikan ke skala semula untuk dilakukan evaluasi metrik dari hasil prediksi tersebut.

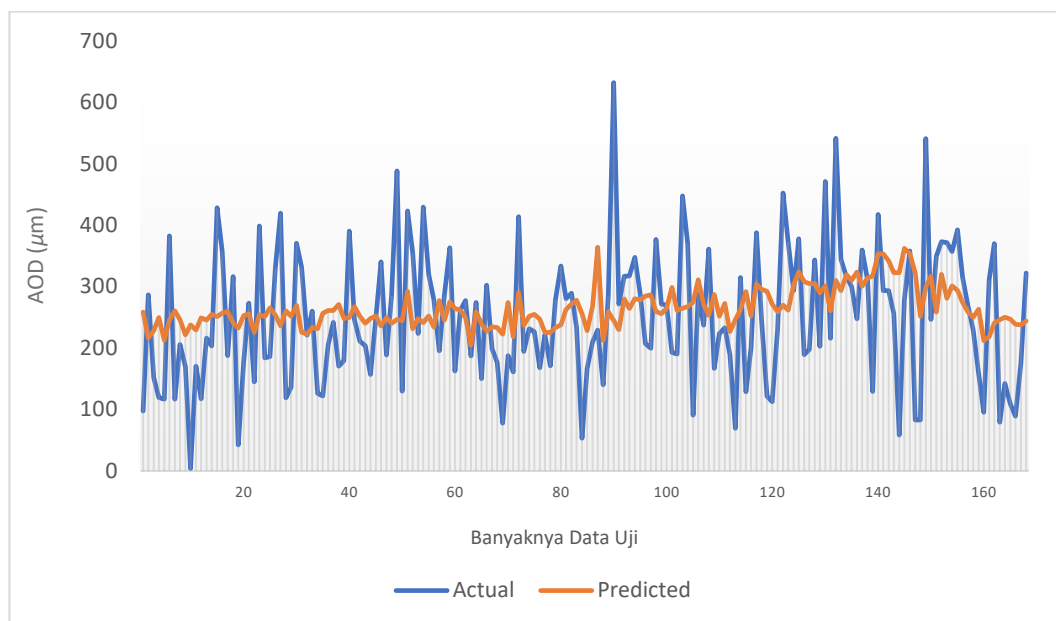
4.4.1.1.1 Model 1

Model 1 menggunakan *split data* dengan perbandingan 75% sebagai data *train* dan 25% sebagai data *test*. Berdasarkan dari *split data* tersebut, dari total 669 baris data, data yang di *training* sebanyak 501 baris data dan 168 baris data untuk di uji. Setelah dilakukan pengujian didapatkanlah hasil prediksi yang ditampilkan pada tabel 4.9 dan gambar 4.8. Pada gambar 4.8, nilai sebenarnya berupa garis berwarna biru dan hasil prediksi berwarna oranye.

Selanjutnya, dilakukan evaluasi pada model. Hasil evaluasi metrik dari pengujian model 1 ditampilkan pada tabel 4.10. yang mencakup *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE).

Tabel 4. 9 Perbandingan Hasil Prediksi Model 1 Normalisasi

No.	<i>Actual</i> AOD (μm)	<i>Predicted</i> AOD (μm)
1.	97,622	258,3493
2.	286,521	216,3532
3.	151,521	228,392
...
166.	89,437	238,6558
167.	174,716	237,8335
168	322	243,7031



Gambar 4. 8 Perbandingan Hasil Prediksi Model 1 Normalisasi

Tabel 4. 10 Hasil Evaluasi Model 1 Normalisasi

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
1 Normalisasi	86,82527641136865	11417,131300788868	106,85097706988395

4.4.1.1.2 Model 2

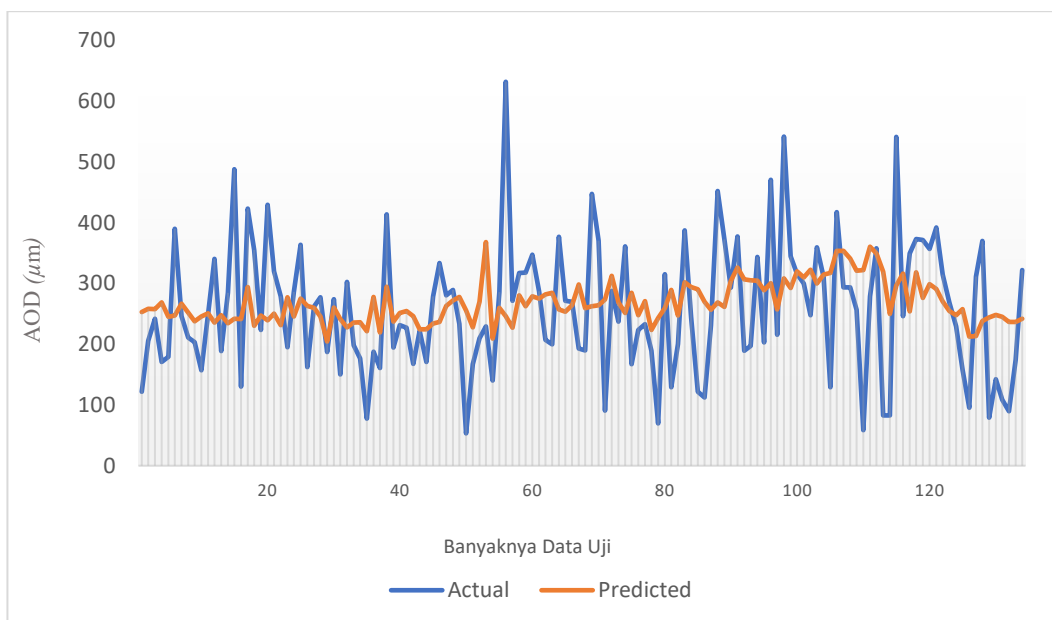
Model 2 menggunakan *split data* dengan rasio 80% data *train* dan 20% data *test*. Keseluruhan total 669 baris data, 535 baris data menjadi *train data* dan 134 baris data menjadi *test data*. Selanjutnya, dilakukan proses prediksi pada model. Perbandingan antara hasil prediksi dan nilai sebenarnya tervisualisasi pada tabel 4.11 dan gambar 4.9. Pada gambar 4.9, nilai sebenarnya berupa garis berwarna biru dan hasil prediksi berwarna oranye.

Langkah berikutnya, model tersebut akan diuji menggunakan metrik evaluasi yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan

Root Mean Squared Error (RMSE). Hasil dari pengujian model ini ditampilkan pada tabel 4.12.

Tabel 4. 11 Perbandingan Hasil Prediksi Model 2 Normalisasi

No.	Actual AOD (μm)	Predicted AOD (μm)
1.	121,872	253,0332
2.	205,506	258,0536
3.	241,983	257,9651
...
132.	89,437	236,7291
133.	174,716	236,7908
134.	322	241,9593



Gambar 4. 9 Perbandingan Hasil Prediksi Model 2 Normalisasi

Tabel 4. 12 Hasil Evaluasi Model 2 Normalisasi

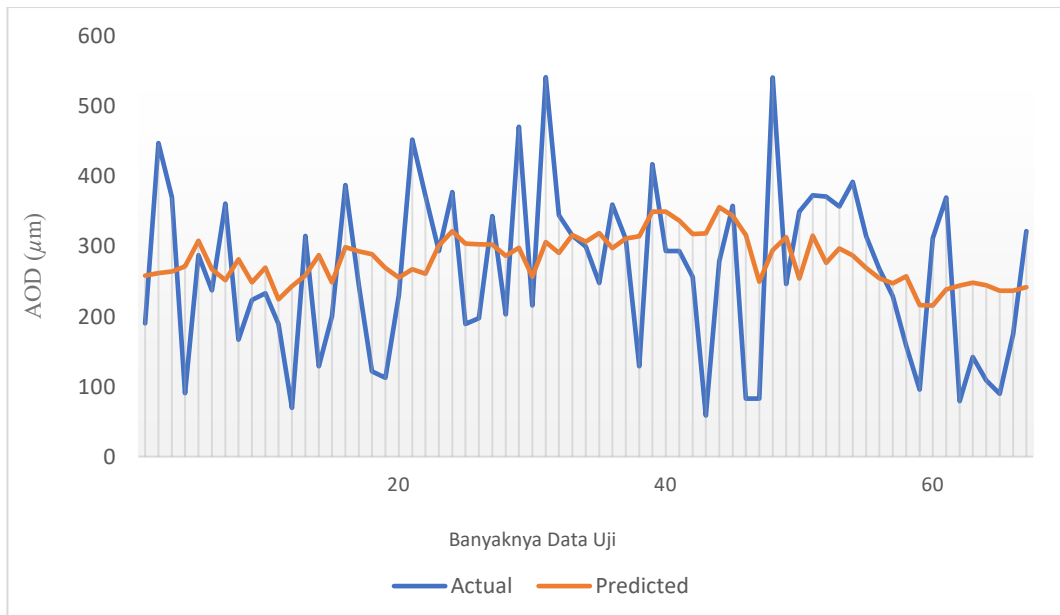
Model	Metrik Evaluasi		
	MAE	MSE	RMSE
2 Normalisasi	83,83957904259945	11140,332314620859	105,5477726653711

4.4.1.1.3 Model 3

Pada model 3 *split data* diterapkan dengan rasio 90% *train data* dan 10% *test data*. Dari total 669 baris data, 541 baris data digunakan untuk proses *training model* dan 67 baris data untuk menguji model. Setelah dilakukan pengujian didapatkanlah hasil prediksi yang selanjutnya dapat dibandingkan pada tabel 4.13 dan gambar 4.10. Pada gambar 4.10, nilai sebenarnya berupa garis berwarna biru dan hasil prediksi berwarna oranye.

Tabel 4. 13 Perbandingan Hasil Prediksi Model 3 Normalisasi

No.	Actual AOD (μm)	Predicted AOD (μm)
1.	190,017	258,1551
2.	447,612	261,7568
3.	370,447	264,3312
...
65.	89,437	236,9599
66.	174,716	236,9818
67.	322	241,9552



Gambar 4. 10 Perbandingan Hasil Prediksi Model 3 Normalisasi

Setelah itu, menggunakan 3 metrik evaluasi yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE) untuk mengevaluasi model. Hasil evaluasi di muat dalam tabel 4.14.

Tabel 4. 14 Hasil Evaluasi Model 3 Normalisasi

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
3 Normalisasi	93,99779032292813	12929,743349831315	113,70902932411003

4.4.1.2 Tanpa Normalisasi

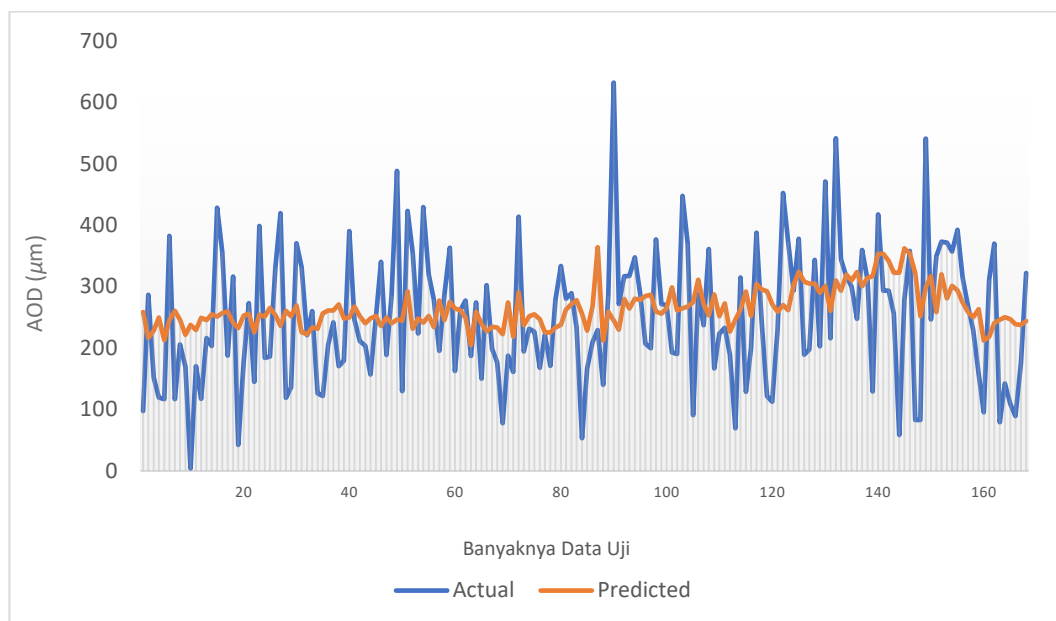
Pengujian dilakukan dengan melewati tahap normalisasi data. Pengujian ini juga dilakukan menggunakan tiga model uji. Ketiga model ini ditentukan berdasarkan *split data* yang dilakukan di tahap *preprocessing*.

4.4.1.2.1 Model 1

Model 1 menggunakan *split data* dengan perbandingan 75% *train data* dan 25% *test data*. Keseluruhan total 669 baris data, data *train* 501 baris data dan data *test* 168 baris data. Kemudian dilakukan pelatihan dan pengujian model sehingga mendapatkan hasil prediksi. Hasil perbandingan antara hasil prediksi dan nilai sebenarnya ditampilkan pada tabel 4.15 dan gambar 4.11. Pada gambar 4.11, nilai sebenarnya berupa garis dengan warna biru dan hasil prediksi warna oranye.

Tabel 4. 15 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi

No.	<i>Actual AOD</i> (μm)	<i>Predicted AOD</i> (μm)
1.	97,622	258,3493
2.	286,521	216,3532
3.	151,521	228,392
...
166.	89,437	238,6558
167.	174,716	237,8335
168.	322	243,7031



Gambar 4. 11 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi

Tahapan selanjutnya, model akan diuji menggunakan metrik evaluasi yaitu MAE, MSE dan RMSE Hasil dari pengujian model 1 ditampilkan pada tabel 4.16.

Tabel 4. 16 Hasil Evaluasi Model 1 Tanpa Normalisasi

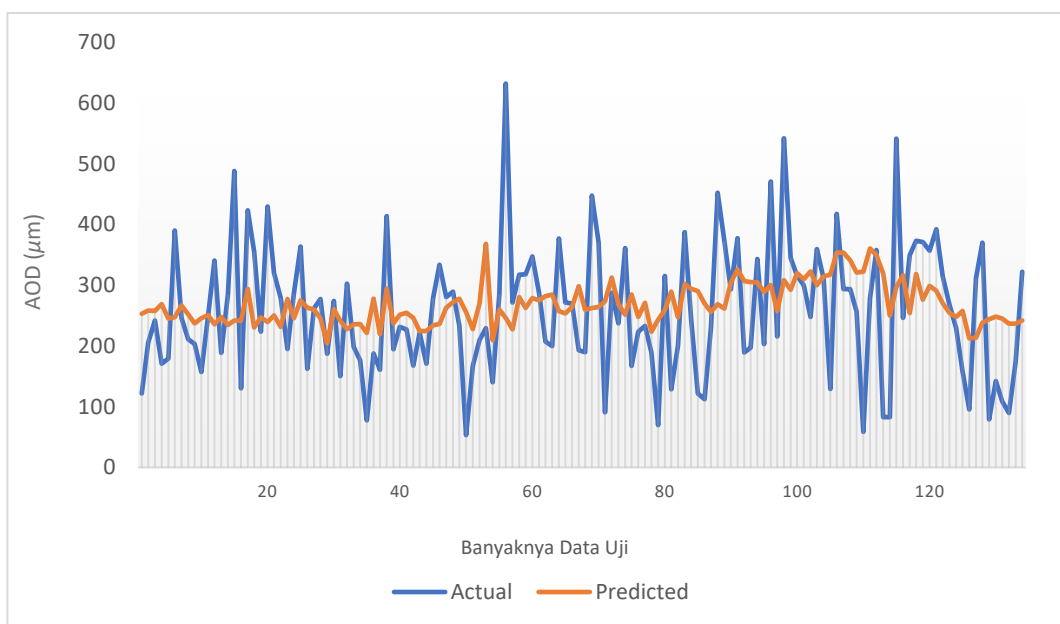
Model	Metrik Evaluasi		
	MAE	MSE	RMSE
1 Tanpa Normalisasi	86,82527641136866	11417,131300788871	106,85097706988398

4.4.1.2.2 Model 2

Pada model 2, rasio pembagian data adalah 80% *train data* dan 20% *test data*. Dari total 669 baris data, sebanyak 535 baris data digunakan untuk *training data* dan sisanya sebanyak 134 baris data digunakan untuk *test data*. Selanjutnya model dilatih dengan data *training* dan diuji dengan data *testing* untuk mendapatkan hasil prediksi. Hasil prediksi dan nilai sebenarnya dari model ini dimuat dalam tabel 4.17 dan gambar 4.12. Pada gambar 4.12, nilai sebenarnya berupa garis berwarna biru dan hasil prediksi berwarna oranye.

Tabel 4. 17 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi

No.	Actual AOD (μm)	Predicted AOD (μm)
1.	121,872	253,0332
2.	205,506	258,0536
3.	241,983	257,9651
...
132.	89,437	236,7291
133.	174,716	236,7908
134.	322	241,9593



Gambar 4. 12 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi

Setelah itu dilakukan evaluasi kinerja model seperti pada tabel 4.18 yang menggunakan 3 metrik evaluasi yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE).

Tabel 4. 18 Hasil Evaluasi Model 2 Tanpa Normalisasi

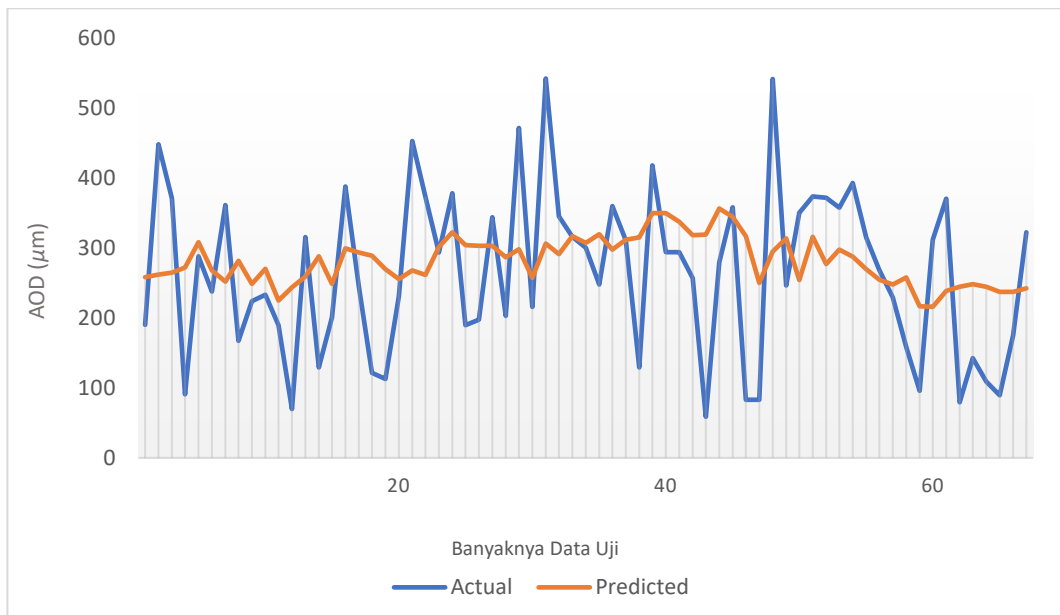
Model	Metrik Evaluasi		
	MAE	MSE	RMSE
2 Tanpa Normalisasi	83,83957904259947	11140,332314620859	105,5477726653711

4.4.1.2.3 Model 3

Model 3 *split data* menerapkan dengan rasio 90% *train data* dan 10% *test data*. Berdasarkan dari *split data* tersebut, total terdapat 669 baris data yang di bagi menjadi data *training* sejumlah 541 baris data dan sejumlah 67 data *testing*. Hasil dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel 4.19 dan gambar 4.13. Di gambar 4.13, nilai sebenarnya berupa garis berwarna biru dan hasil prediksi berwarna oranye.

Tabel 4. 19 Perbandingan Hasil Prediksi Model 3 Tanpa Normalisasi

NO.	Actual AOD (μm)	Predicted AOD (μm)
1.	190,017	258,1551
2.	447,612	261,7568
3.	370,447	264,3312
...
65.	89,437	236,9599
66.	174,716	236,9818
67.	322	241,9552



Gambar 4. 13 Perbandingan Hasil Prediksi Model 3 Tanpa Normalisasi

Tahap berikutnya, dilakukan evaluasi model menggunakan metrik evaluasi dengan menggunakan *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE) yang termuat dalam tabel 4.20.

Tabel 4. 20 Hasil Evaluasi Model 3 Tanpa Normalisasi

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
3 Tanpa Normalisasi	93,99779032292811	12929,743349831317	113,70902932411003

4.4.2 Data Tahunan

4.4.2.1 Tahun 2019

4.4.2.1.1 Dengan Normalisasi

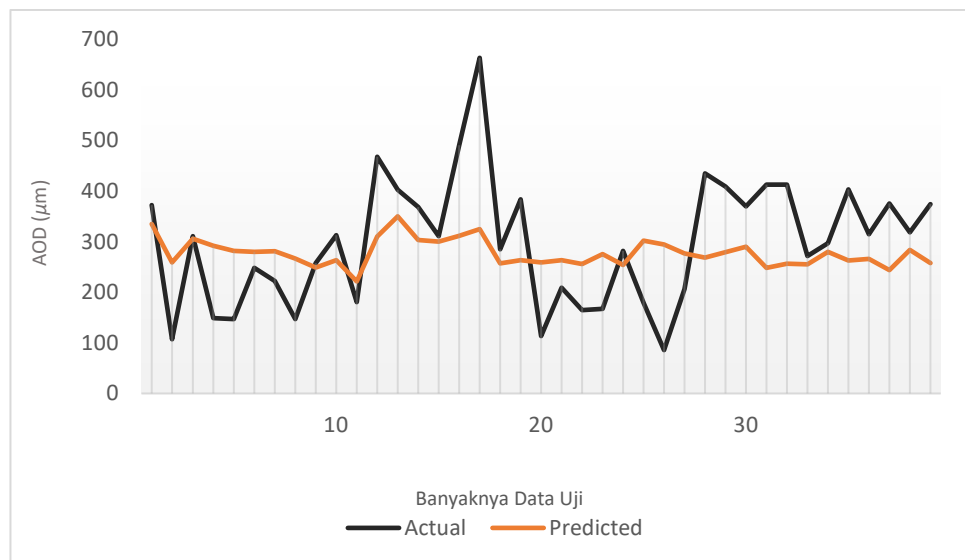
Pada pengujian ini ditambahkan normalisasi data pada proses *preprocessing*. Terdapat tiga model yang akan diujikan. Perbedaan antar tiga model tersebut didasarkan pada perbandingan rasio *split data*. Setelah dilakukan implementasi metode *linear regression* didapatkanlah hasil prediksi. Kemudian hasil prediksi tersebut dikembalikan ke skala semula untuk dilakukan evaluasi metrik dari hasil prediksi tersebut.

A. Model 1

Model 1 menerapkan *split data* dengan perbandingan 75% data *train* dan 25% data *test*. Berdasarkan dari *split data* tersebut, total terdapat 156 baris data yang di bagi menjadi data *training* sebanyak 117 baris data dan data *testing* sebanyak 39 data. Dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya pada tabel 4.21 dan gambar 4.14. Pada gambar 4.14, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna oranye.

Tabel 4. 21 Perbandingan Hasil Prediksi Model 1 Normalisasi 2019

No.	Actual AOD (μm)	Predicted AOD (μm)
1.	372,417	335,0782
2.	107,004	258,9497
3.	310,648	305,8636
...
37.	375,351	243,6562
38.	318,217	283,5945
39.	374,031	257,769



Gambar 4. 14 Perbandingan Hasil Prediksi Model 1 Normalisasi 2019

Tahapan selanjutnya mengevaluasi model menggunakan metrik evaluasi dengan menggunakan *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Hasil metrik evaluasi termuat dalam tabel 4.22.

Tabel 4. 22 Hasil Evaluasi Model 1 Normalisasi 2019

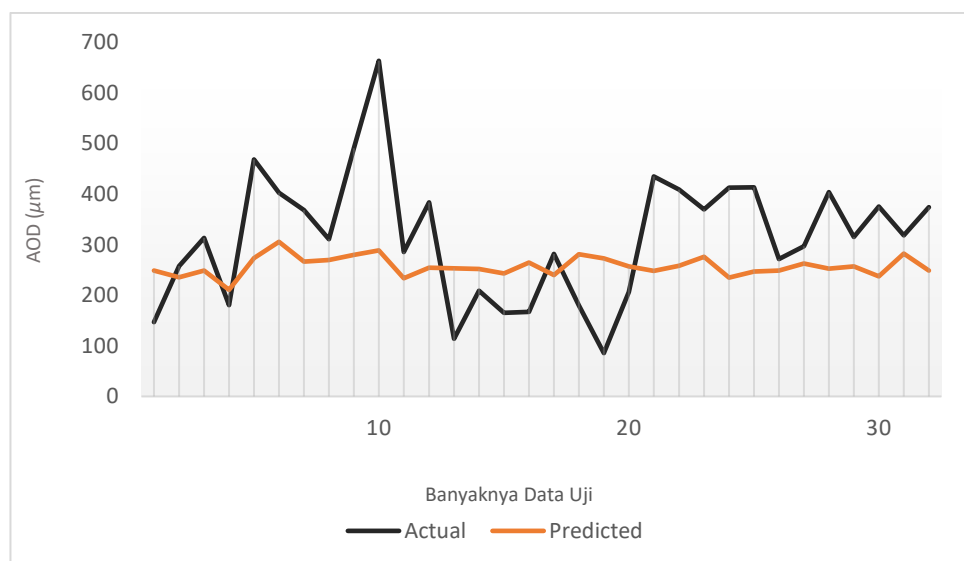
Model	Metrik Evaluasi		
	MAE	MSE	RMSE
1 Normalisasi 2019	96,4761913320 8842	13977,48276501 2692	118,2264046861 4738

B. Model 2

Model 2 menggunakan *split data* dengan rasio 80% *train data* dan 20% *test test*. Dari total 156 baris data, 124 baris data menjadi data *train* dan 32 baris data menjadi data *test*. Kemudian dilakukan pelatihan dan pengujian model sehingga mendapatkan hasil prediksi. Perbandingan antara hasil prediksi dan nilai sebenarnya ditampilkan di tabel 4.23 dan gambar 4.15. Di gambar 4.15, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna oranye

Tabel 4. 23 Perbandingan Hasil Prediksi Model 2 Normalisasi 2019

No.	Actual AOD (μm)	Predicted AOD (μm)
1.	146,733	248,5817
2.	257,434	235,4167
3.	313,026	248,9564
..
30.	375,351	237,3282
31.	318,217	282,0999
32.	374,031	248,8199



Gambar 4. 15 Perbandingan Hasil Prediksi Model 1 Normalisasi 2019

Model ini selanjutnya diuji menggunakan metrik evaluasi yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Hasil dari pengujian model 1 ditampilkan pada tabel 4.24.

Tabel 4. 24 Hasil Evaluasi Model 1 Normalisasi 2019

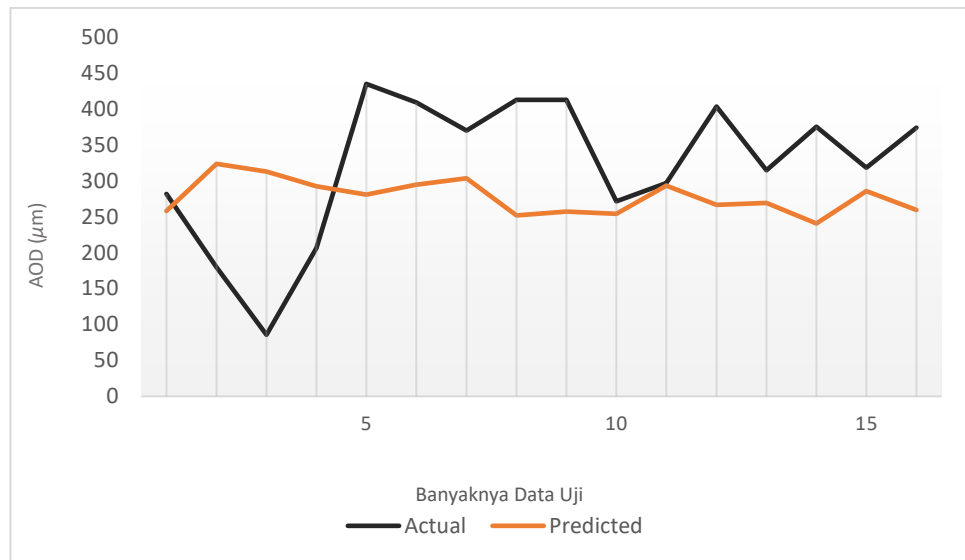
Model	Metrik Evaluasi		
	MAE	MSE	RMSE
2 Normalisasi 2019	109,34379176846 039	17363,30202679 657	131,7698828518 739

C. Model 3

Pada model 3 *split data* diterapkan dengan rasio 90% *train data* dan 10% *test data*. Dari total 156 baris data, 140 baris data diproses menjadi *training model* dan 16 baris data sebagai pengujian model. Setelah dilakukan pengujian didapatkanlah hasil prediksi dalam bentuk yang dinormalisasi. Selanjutnya, tabel 4.25 dan gambar 4.16 ditampilkan perbandingan nilai dari hasil prediksi yang telah diubah ke skala semula dan nilai sebenarnya. Pada gambar 4.16, hasil prediksi yang berupa garis oranye dan nilai sebenarnya yang berwarna hitam.

Tabel 4. 25 Perbandingan Hasil Prediksi Model 3 Normalisasi 2019

No.	<i>Actual AOD</i> (μm)	<i>Predicted AOD</i> (μm)
1.	281,532	258,3201
2.	179,504	323,702
3.	85,698	313,007
...
14.	375,351	240,7091
15.	318,217	285,8622
16.	374,031	259,357



Gambar 4. 16 Perbandingan Hasil Prediksi Model 3 Normalisasi 2019

Tahapan selanjutnya dilakukan evaluasi kinerja model dengan menggunakan 3 metrik evaluasi yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Hasil evaluasi di muat dalam tabel 4.26.

Tabel 4. 26 Hasil Evaluasi Model 3 Normalisasi 2019

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
3 Normalisasi 2019	101,08566742 5938	14082,78130684 1593	118,670894944 1336

4.4.2.1.2 Tanpa Normalisasi

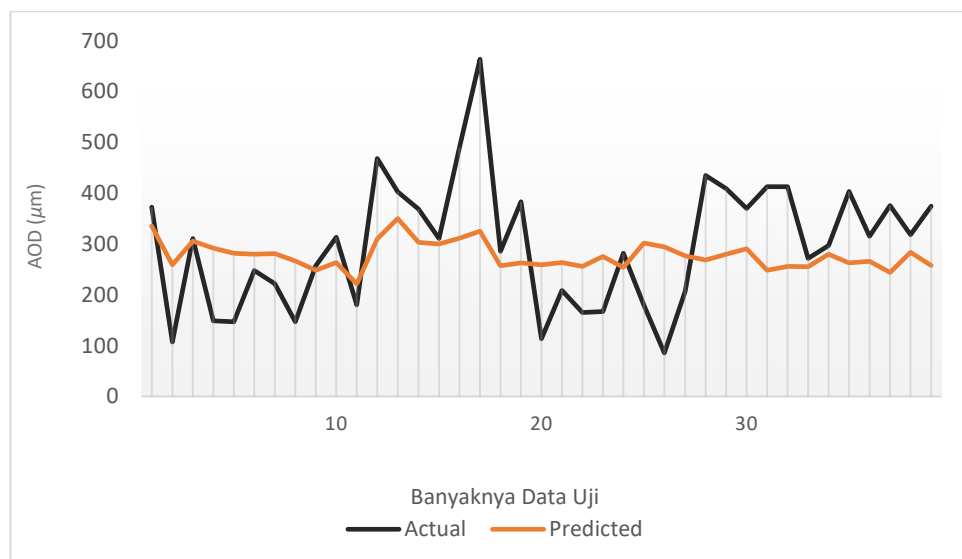
Pengujian dilakukan dengan melewati tahap normalisasi data. Pengujian ini juga dilakukan menggunakan tiga model uji. Ketiga model ini ditentukan berdasarkan *split data* yang dilakukan di tahap *preprocessing*.

A. Model 1

Model 1 menggunakan *split data* dengan perbandingan 75% *train data* dan 25% *test data*. Keseluruhan total 156 baris data, 117 baris data *train* dan 39 baris data *test*. Kemudian dilakukan pelatihan dan pengujian model sehingga mendapatkan hasil prediksi. Perbedaan antara prediksi AOD dan aktual AOD ditampilkan di tabel 4.27 dan gambar 4.17. Di gambar 4.17, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna oranye.

Tabel 4. 27 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi

No.	Actual AOD (μm)	Predicted AOD (μm)
1.	372,417	335,0782
2.	107,004	258,9497
3.	310,648	305,8636
...
37.	375,351	243,6562
38.	318,217	283,5945
39.	374,031	257,769



Gambar 4. 17 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi 2019

Tahapan selanjutnya dilakukan evaluasi kinerja model dengan menggunakan 3 metrik evaluasi yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Hasil evaluasi di muat dalam tabel 4.28.

Tabel 4. 28 Hasil Evaluasi Model 1 Tanpa Normalisasi 2019

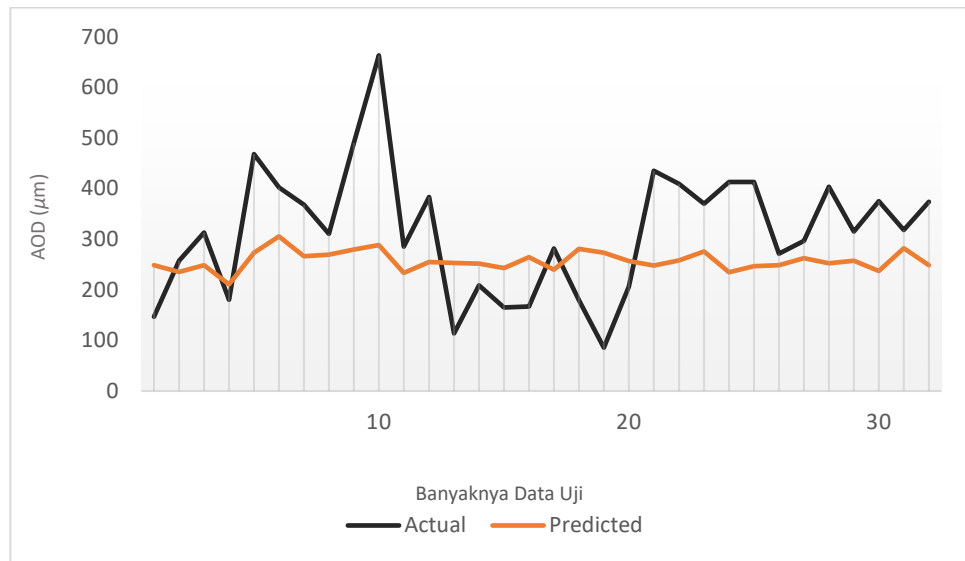
Model	Metrik Evaluasi		
	MAE	MSE	RMSE
1 Tanpa Normalisasi 2019	96,47619133208 848	13977,48276501 27	118,22640468614 742

B. Model 2

Model 2 menggunakan *split data* dengan rasio 80% *train data* dan 20% *test data*. Dari total 156 baris data, data *train* 124 baris data dan data *test* 32 baris data. Kemudian dilakukan pelatihan dan pengujian model sehingga mendapatkan hasil prediksi. Rasio hasil prediksi dan nilai aktual ditampilkan di tabel 4.29 dan gambar 4.18. Di gambar 4.18, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna oranye.

Tabel 4. 29 Peerbandingan Hasil Prediksi Model 2 Tanpa Normalisasi 2019

No.	<i>Actual AOD</i> (μm)	<i>Predicted AOD</i> (μm)
1.	146,733	248,5817
2.	257,434	235,4167
3.	313,026	248,9564
...
30.	375,351	237,3282
31.	318,217	282,0999
32.	374,031	248,8199



Gambar 4. 18 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi 2019

Model ini selanjutnya diuji menggunakan metrik evaluasi yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Hasil dari pengujian model 1 ditampilkan pada tabel 4.30.

Tabel 4. 30 Hasil Evaluasi Model 2 Tanpa Normalisasi

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
2 Tanpa Normalisasi 2019	109,34379176846011	17363,302026796486	131,7698828587358

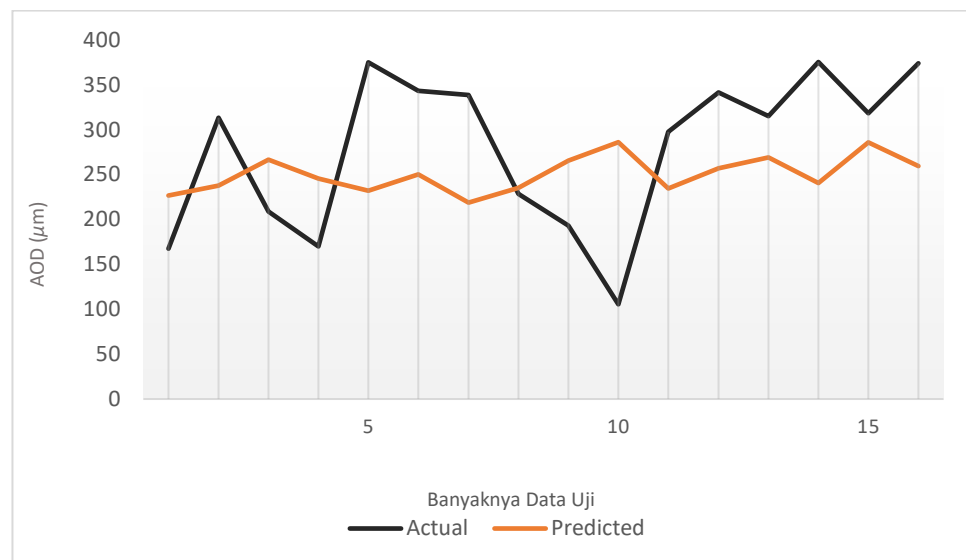
C. Model 3

Model 3 menerapkan rasio *split data* 90% *train data* dan 10% *test data*. Berdasarkan dari *split data* tersebut, total terdapat 156 baris data yang di bagi menjadi data *training* sejumlah 140 baris data dan data *testing* sejumlah 16 data. Hasil dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel 4.31 dan

gambar 4.19. Di gambar 4.19, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna oranye

Tabel 4. 31 Perbandingan Hasil Prediksi Model 3 Tanpa Normalisasi 2019

No.	Actual AOD (μm)	Predicted AOD (μm)
1.	281,532	258,3201
2.	179,504	323,702
3.	85,698	313,007
...
14.	375,351	240,7091
15.	318,217	285,8622
16.	374,031	259,357



Gambar 4. 19 Perbandingan Hasil Evaluasi Model 3 Tanpa Normalisasi 2019

Model ini selanjutnya diuji menggunakan metrik evaluasi yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Hasil dari pengujian model 1 ditampilkan pada tabel 4.32.

Tabel 4. 32 Hasil Evaluasi Model 3 Tanpa Normalisasi 2019

Tahun	Metrik Evaluasi		
	MAE	MSE	RMSE
3 Tanpa Normalisasi 2019	101,08566742593803	14082,7813068416	14082,7813068416

4.4.2.2 Tahun 2020

4.4.2.2.1 Dengan Normalisasi

Pada pengujian ini ditambahkan normalisasi data pada proses *preprocessing*. Terdapat tiga model yang akan diujikan. Perbedaan antar tiga model tersebut didasarkan pada perbandingan rasio *split data*. Setelah dilakukan implementasi metode *linear regression* didapatkanlah hasil prediksi. Kemudian hasil prediksi tersebut dikembalikan ke skala semula untuk dilakukan evaluasi metrik dari hasil prediksi tersebut.

A. Model 1

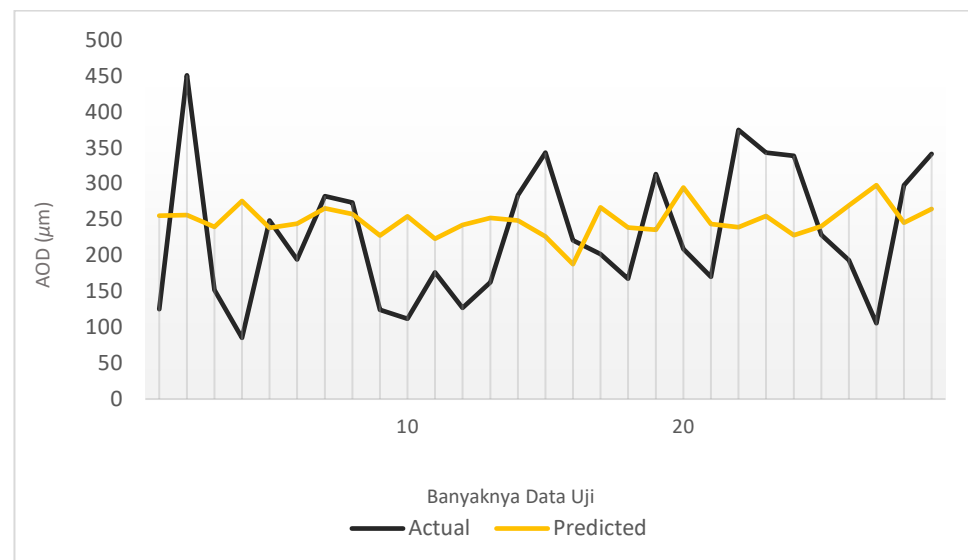
Model 1 menggunakan *split data* dengan perbandingan 75% *train data* dan 25% *test data*. Keseluruhan total 115 baris data, 86 baris data *train* dan 29 baris data *test*. Kemudian dilakukan pelatihan dan pengujian model sehingga mendapatkan hasil prediksi. Perbandingan antara hasil prediksi dan nilai sebenarnya ditampilkan di tabel 4.33 dan gambar 4.20. Di gambar 4.20, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna kuning.

Selanjutnya, dilakukan evaluasi pada model. Hasil evaluasi metrik dari pengujian model 1 ditampilkan pada tabel 4.34. yang mencakup *Mean*

Absolute Error (MAE), Mean Squared Error (MSE), dan Root Mean Squared Error (RMSE).

Tabel 4. 33 Perbandingan Hasil Prediksi Model 1 Normalisasi 2020

No.	Actual AOD (μm)	Predicted AOD (μm)
1.	124,938	255,1634
2.	450,694	256,2633
3.	151,883	239,6803
...
27.	105,496	297,748
28.	297,682	245,6121
29.	341,347	264,4351



Gambar 4. 20 Perbandingan Hasil Prediksi Model 1 Normalisasi 2020

Tabel 4. 34 Hasil Evaluasi Model 1 Normalisasi 2020

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
1 Normalisasi 2020	86,08453440384 996	10020,486182971 323	100,10237850806 205

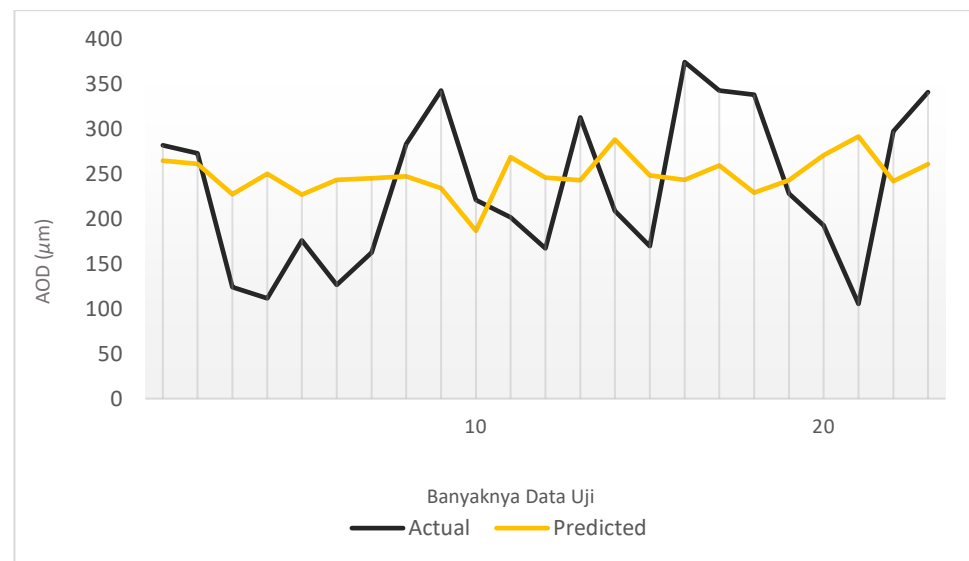
B. Model 2

Model 2 menerapkan *split data* dengan rasio 80% *train data* dan 20% *test data*. Keseluruhan total 115 baris data, 92 baris data menjadi data

train dan 23 baris data menjadi data *test*. Selanjutnya, dilakukan proses prediksi pada pada model. Dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel 4.35 dan gambar 4.21. Di gambar 4.21, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna kuning.

Tabel 4. 35 Perbandingan Hasil Prediksi Model 2 Normalisasi 2020

No.	Actual AOD (μm)	Predicted AOD (μm)
1.	282,269	264,9506
2.	273,402	261,3303
3.	124,086	227,4722
...
21.	105,496	291,9805
22.	297,682	242,3431
23.	341,347	261,2175



Gambar 4. 21 Perbandingan Hasil Prediksi Model 2 Normalisasi 2020

Langkah berikutnya, model tersebut akan diuji menggunakan metrik evaluasi yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Hasil dari pengujian model ini ditampilkan pada tabel 4.36.

Tabel 4. 36 Hasil Evaluasi Model 2 Normalisasi 2020

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
2 Normalisasi 2020	78,84943592611091	7934,003612790981	89,07302404651467

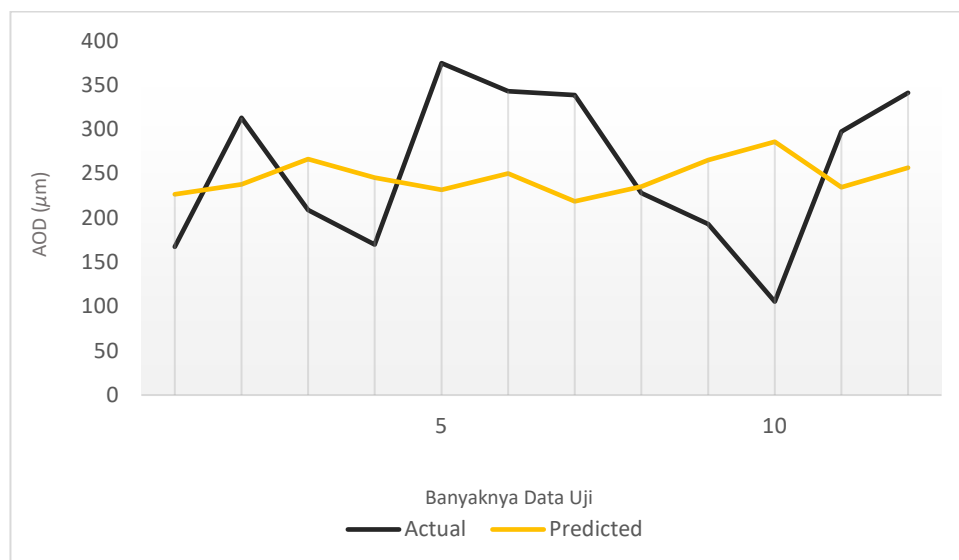
C. Model 3

Pada model 3, *split data* diterapkan dengan rasio 90% *train data* dan 10% *test data*. Dari total 115 baris data, 103 baris data digunakan untuk proses *training model* dan 12 baris data untuk menguji model. Hasil dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel 4.37 dan gambar 4.22. di gambar 4.22, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna kuning.

Tahapan selanjutnya mengevaluasi model menggunakan metrik evaluasi dengan menggunakan *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Hasil metrik evaluasi termuat dalam tabel 4.38.

Tabel 4. 37 Perbandingan Hasil Prediksi Model 3 Normalisasi 2020

No.	<i>Actual AOD</i> (μm)	<i>Predicted AOD</i> (μm)
1.	167,448	226,5675
2.	313,155	237,7534
3.	209,008	266,4426
...
10.	105,496	286,0587
11.	297,682	234,5572
12.	341,347	256,7916



Gambar 4. 22 Perbandingan Hasil Prediksi Model 3 Normalisasi 2020

Tabel 4. 38 Perbandingan Hasil Prediksi Model 3 Normalisasi 2020

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
3 Normalisasi 2020	85,90267303245 933	9220,032129764 912	96,02099837933 842

4.4.2.2.2 Tanpa Normalisasi

Pengujian dilakukan dengan melewati tahap normalisasi data. Pengujian ini juga dilakukan menggunakan tiga model uji. Ketiga model ini ditentukan berdasarkan *split data* yang dilakukan di tahap *preprocessing*.

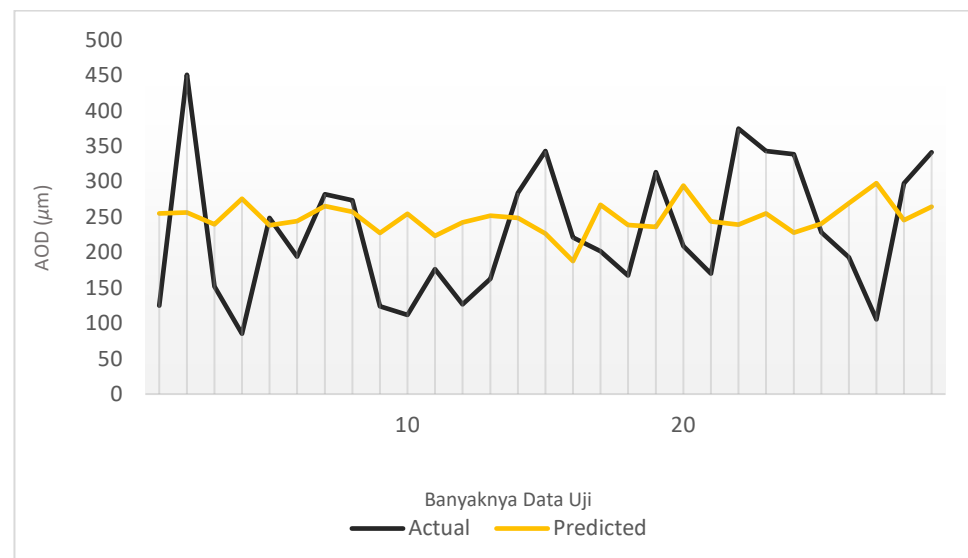
A. Model 1

Model 1 menggunakan *split data* dengan perbandingan 75% sebagai data *train* dan 25% sebagai data *test*. Berdasarkan dari split data tersebut, dari total 115 baris data, data yang di *training* sebanyak 86 baris data dan 29 baris data untuk di uji. Dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel

4.39 dan gambar 4.23. Di gambar 4.23, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna kuning.

Tabel 4. 39 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi 2020

No.	Actual AOD (μm)	Predicted AOD (μm)
1.	124,938	255,1634
2.	450,694	256,2633
3.	151,883	239,6803
...
27.	105,496	297,748
28.	297,682	245,6121
29.	341,347	264,4351



Gambar 4. 23 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi 2020

Selanjutnya, dilakukan evaluasi pada model. Hasil evaluasi metrik dari pengujian model 1 ditampilkan pada tabel 4.40. yang mencakup *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE).

Tabel 4. 40 Hasil Evaluasi Model 1 Tanpa Normalisasi 2020

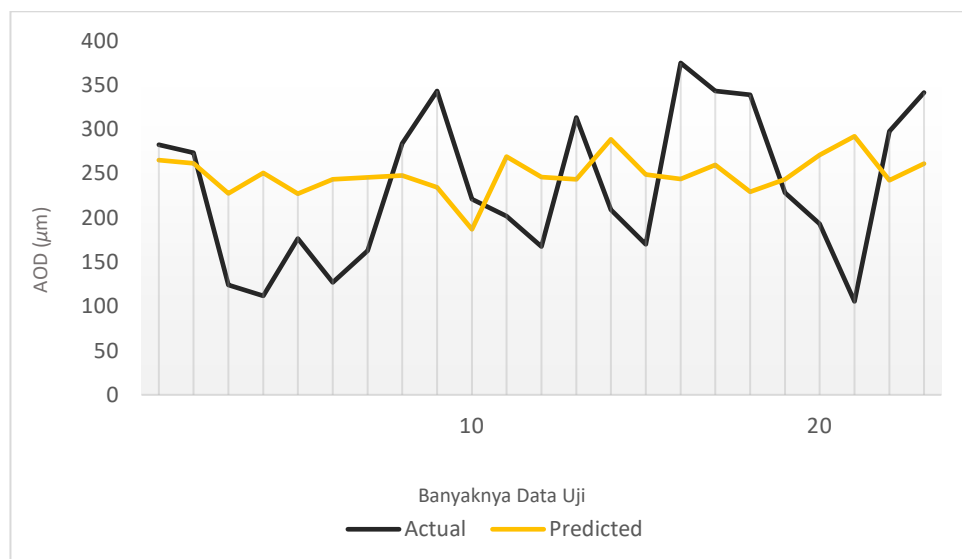
Model	Metrik Evaluasi		
	MAE	MSE	RMSE
1 Tanpa Normalisasi 2020	86,0845344038 5001	10020,4861829 71335	100,102378508 06212

B. Model 2

Model 2, *split data* menggunakan rasio 80% *train data* dan 20% *test data*. Dari total 115 baris data, 92 baris data menjadi data *train* dan 23 baris data menjadi data *test*. Selanjutnya, dilakukan proses prediksi pada pada model. Dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel 4.41 dan gambar 4.24. Di gambar 4.24, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna kuning.

Tabel 4. 41 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi 2020

No.	Actual AOD (μm)	Predicted AOD (μm)
1.	282,269	264,9506
2.	273,402	261,3303
3.	124,086	227,4722
...
21.	105,496	291,9805
22.	297,682	242,3431
23.	341,347	261,2175



Gambar 4. 24 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi

Langkah berikutnya, model tersebut akan diuji menggunakan metrik evaluasi yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Hasil dari pengujian model ini ditampilkan pada tabel 4.42.

Tabel 4. 42 Hasil Evaluasi Model 2 Tanpa Normalisasi 2020

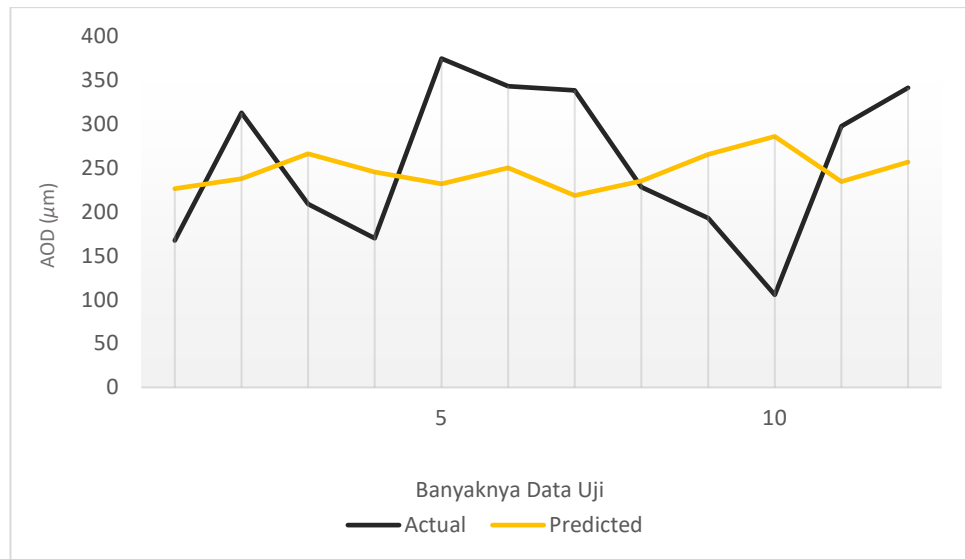
Model	Metrik Evaluasi		
	MAE	MSE	RMSE
2 Tanpa Normalisasi 2020	78,8494359261 1101	7934,00361279 1004	89,0730240465 148

C. Model 3

Pada model 3 *split data* diterapkan dengan rasio 90% *train data* dan 10% *test data*. Dari total 115 baris data, 103 baris data untuk proses *training model* dan 12 baris data untuk pengujian. Hasil dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel 4.43 dan gambar 4.25. Pada gambar 4.25, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna kuning.

Tabel 4. 43 Perbandingan Hasil Prediksi Model 3 Tanpa Normalisasi 2020

No.	<i>Actual AOD</i> (μm)	<i>Predicted AOD</i> (μm)
1.	167,448	226,5675
2.	313,155	237,7534
3.	209,008	266,4426
...
10.	105,496	286,0587
11.	297,682	234,5572
12.	341,347	256,7916



Gambar 4. 25 Perbandingan Hasil Prediksi Model 3 Tanpa Normalisasi

Tahapan selanjutnya mengevaluasi model menggunakan metrik evaluasi dengan menggunakan *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Hasil metrik evaluasi termuat dalam tabel 4.44.

Tabel 4. 44 Hasil Evaluasi Model 3 Tanpa Normalisasi 2020

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
3 Tanpa Normalisasi 2020	85,9026730324 5948	9220,03212976 4943	96,0209983793 3858

4.4.2.3. Tahun 2021

4.4.2.3.1 Dengan Normalisasi

Pada pengujian ini ditambahkan normalisasi data pada proses *preprocessing*. Terdapat tiga model yang akan diujikan. Perbedaan antar tiga model tersebut didasarkan pada perbandingan rasio *split data*. Setelah dilakukan

implementasi metode *linear regression* didapatkanlah hasil prediksi. Kemudian hasil prediksi tersebut dikembalikan ke skala semula untuk dilakukan evaluasi metrik dari hasil prediksi tersebut.

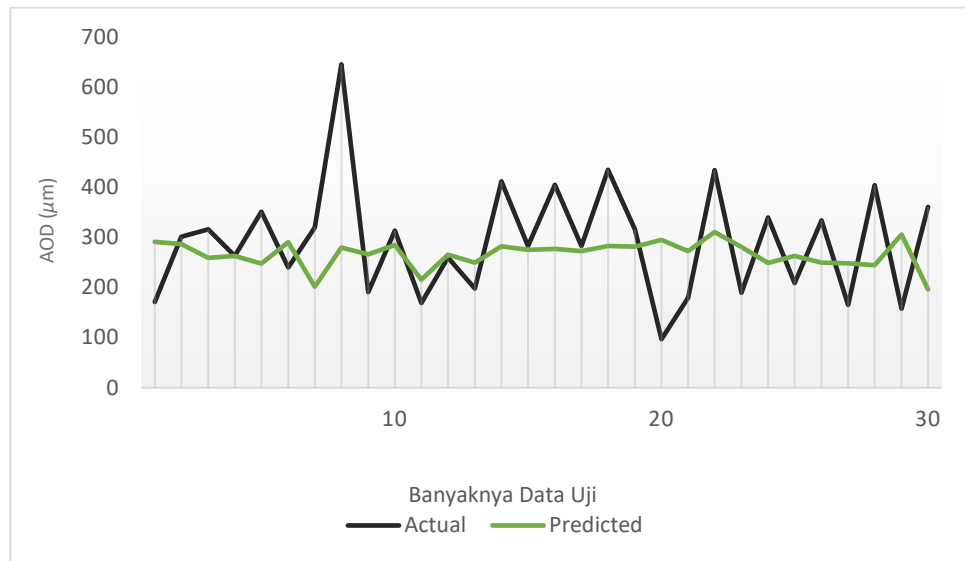
A. Model 1

Model 1 menggunakan *split data* dengan perbandingan 75% sebagai data *train* dan 25% sebagai data *test*. Berdasarkan dari *split data* tersebut, dari total 117 baris data, data yang di *training* sebanyak 87 baris data dan 30 baris data untuk di uji. Dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya pada tabel 4.45 dan gambar 4.26. Pada gambar 4.26, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna hijau.

Selanjutnya, dilakukan evaluasi pada model. Hasil evaluasi metrik dari pengujian model 1 ditampilkan pada tabel 4.46. yang mencakup *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE).

Tabel 4. 45 Perbandingan Hasil Prediksi Model 1 Normalisasi 2021

No.	<i>Actual AOD</i> (μm)	<i>Predicted AOD</i> (μm)
1.	170,862	290,8831
2.	301,477	286,4503
3.	315,978	259,1484
...
28.	403,86	244,2089
29.	157,536	305,9898
30.	360,894	196,0812



Gambar 4. 26 Perbandingan Hasil Prediksi Model 1 Normalisasi 2021

Tabel 4. 46 Hasil Evaluasi Model 1 Normalisasi 2021

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
1 Normalisasi 2021	92,94774237090468	13952,85646891004	118,12220988836113

B. Model 2

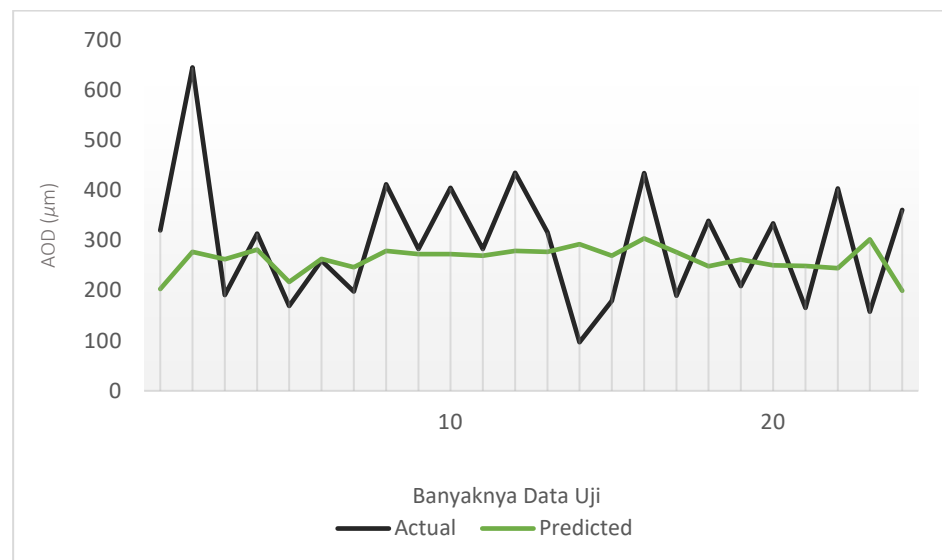
Model 2 menggunakan *split data* dengan rasio 80% *train data* dan 20% *test data*. Dari total 117 baris data, *train data* 93 baris data dan *test data* 24 baris data. Selanjutnya, dilakukan proses prediksi pada pada model. Dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel 4.47 dan gambar 4.27. Pada gambar 4.27, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna hijau.

Langkah berikutnya, model tersebut akan diuji menggunakan metrik evaluasi yaitu *Mean Absolute Error* (MAE), *Mean Squared Error*

(MSE), dan *Root Mean Squared Error* (RMSE). Hasil dari pengujian model ini ditampilkan pada tabel 4.48.

Tabel 4. 47 Perbandingan Hasil Prediksi Model 2 Normalisasi 2021

No.	Actual AOD (μm)	Predicted AOD (μm)
1.	320	202,9788
2.	645,053	276,8627
3.	191,1	262,4847
...
22.	403,86	244,7963
23.	157,536	302,2328
24.	360,894	199,5853



Gambar 4. 27 Perbandingan Hasil Prediksi Model 2 Normalisasi 2021

Tabel 4. 48 Hasil Evaluasi Model 2 Normalisasi 2021

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
2 Normalisasi 2021	102,047262700 77727	16213,7076658 72265	127,333058024 50622

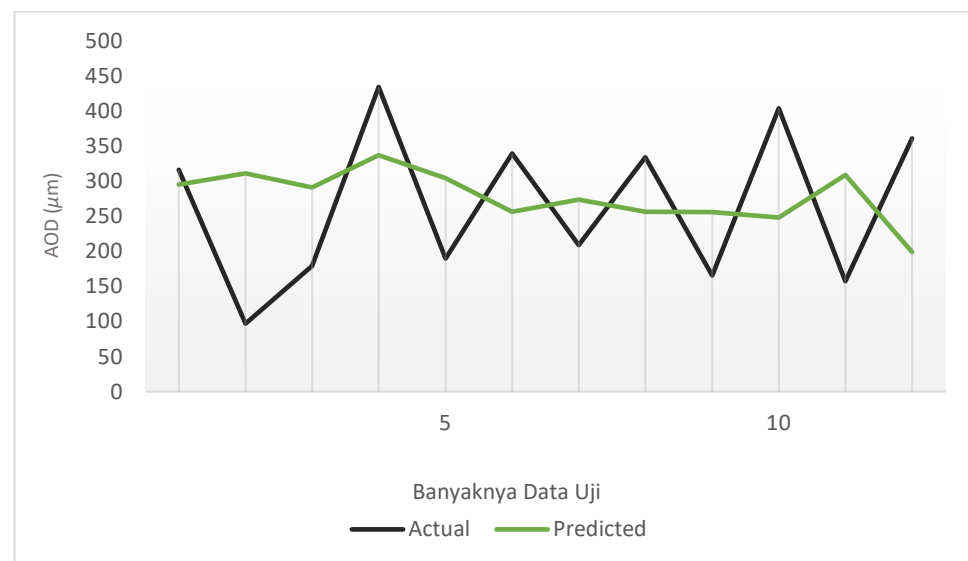
C. Model 3

Pada model 3 *split data* diterapkan dengan rasio 90% *train data* dan 10% *test data*. Dari total 117 baris data, 105 baris data diproses untuk

training model dan 12 baris data untuk pengujian model. Hasil dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel 4.9 dan gambar 4.28. Pada gambar 4.28, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna hijau.

Tabel 4. 49 Perbandingan Hasil Prediksi Model 3 Normalisasi 2021

No.	Actual AOD (μm)	Predicted AOD (μm)
1.	316,011	294,9354
2.	96,914	310,9652
3.	179,248	290,9733
...
10.	403,86	248,3436
11.	157,536	308,7991
12.	360,894	198,9661



Gambar 4. 28 Perbandingan Hasil Prediksi Model 3 Normalisasi 2021

Tahapan selanjutnya mengevaluasi model menggunakan metrik evaluasi dengan menggunakan *Mean Absolute Error* (MAE), *Mean*

Squared Error (MSE), dan *Root Mean Squared Error* (RMSE). Hasil metrik evaluasi termuat dalam tabel 4.50.

Tabel 4. 50 Hasil Evaluasi Model 3 Normalisasi 2021

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
3 Normalisasi 2021	111,91196686907978	14988,532992548746	122,4276643269353

4.4.2.3.2 Tanpa Normalisasi

Pengujian dilakukan dengan melewati tahap normalisasi data. Pengujian ini juga dilakukan menggunakan tiga model uji. Ketiga model ini ditentukan berdasarkan *split data* yang dilakukan di tahap *preprocessing*.

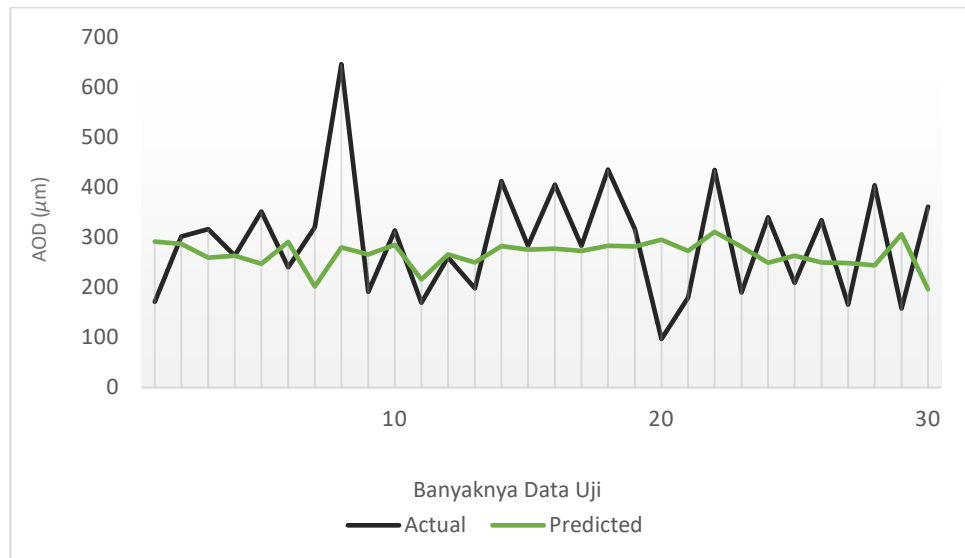
A. Model 1

Model 1 menggunakan *split data* dengan perbandingan 75% sebagai data *train* dan 25% sebagai data *test*. Berdasarkan dari split data tersebut, dari total 117 baris data, data yang di *training* sebanyak 87 baris data dan 30 baris data untuk di uji. Dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel 4.51 dan gambar 4.29. Pada gambar 4.29, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna oranye. Selanjutnya, dilakukan evaluasi pada model. Hasil evaluasi metrik dari pengujian model 1 ditampilkan pada tabel 4.52. yang mencakup *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE).

Tabel 4. 51 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi 2021

No.	<i>Actual AOD</i> (μm)	<i>Predicted AOD</i> (μm)
1.	170,862	290,8831

2.	301,477	286,4503
3.	315,978	259,1484
...
28.	403,86	244,2089
29.	157,536	305,9898
30.	360,894	196,0812



Gambar 4. 29 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi 2021

Tabel 4. 52 Hasil Evaluasi Model 1 Tanpa Normalisasi 2021

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
1 Tanpa Normalisasi 2021	92,94774237090463	13952,856468910024	118,12220988836106

B. Model 2

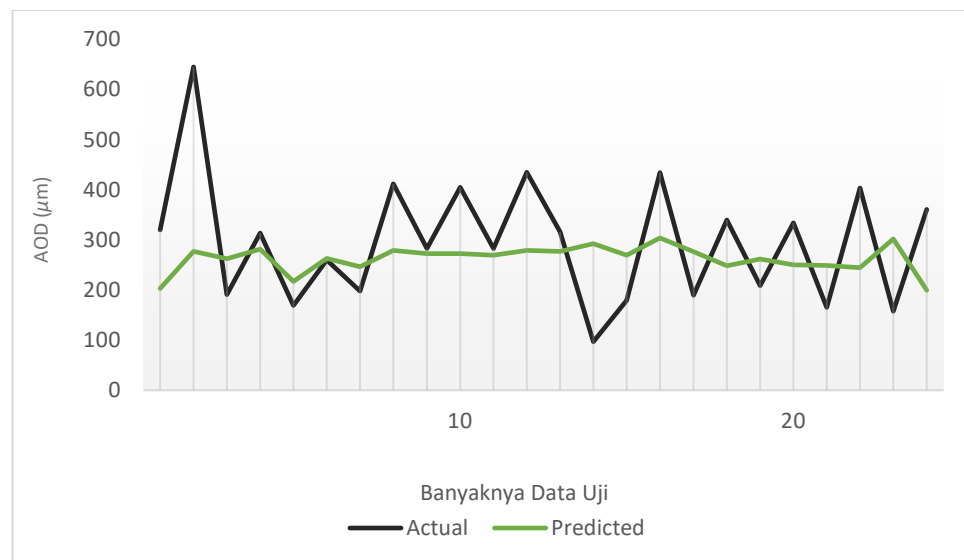
Model 2 menggunakan *split data* dengan rasio pembagian 80% *train data* dan 20% *test data*. Dari total 117 baris data, 93 baris data menjadi data *train* dan 24 baris data menjadi data *test*. Selanjutnya, dilakukan proses prediksi pada pada model. Dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai

sebenarnya di tabel 4.53 dan gambar 4.30. Di gambar 4.30, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna hijau.

Langkah berikutnya, model tersebut akan diuji menggunakan metrik evaluasi yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Hasil dari pengujian model ini ditampilkan pada tabel 4.54.

Tabel 4. 53 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi 2021

No.	Actual AOD (μm)	Predicted AOD (μm)
1.	320	202,9788
2.	645,053	276,8627
3.	191,1	262,4847
...
22.	403,86	244,7963
23.	157,536	302,2328
24.	360,894	199,5853



Gambar 4. 30 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi 2021

Tabel 4. 54 Hasil Evaluasi Model 2 Tanpa Normalisasi 2021

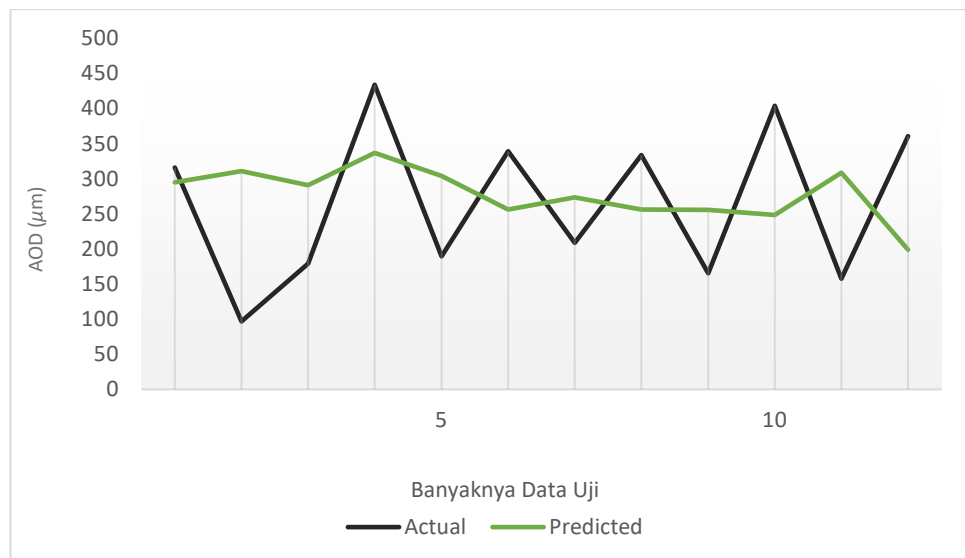
Model	Metrik Evaluasi		
	MAE	MSE	RMSE
2 Tanpa Normalisasi 2021	102,0472627007 7735	16213,70766587 2287	127,333058024 5063

C. Model 3

Pada model 3 *split data* diterapkan dengan rasio 90% *train data* dan 10% *test data*. Dari total 117 baris data 105 baris data digunakan untuk proses *training model* dan 12 baris data untuk menguji model. Hasil dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel 4.55 dan gambar 4.31. Pada gambar 4.31, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna hijau.

Tabel 4. 55 Perbandingan Hasil Prediksi Model 3 Tanpa Normalisasi 2021

No.	<i>Actual AOD</i> (μm)	<i>Predicted AOD</i> (μm)
1.	316,011	294,9354
2.	96,914	310,9652
3.	179,248	290,9733
...
10.	403,86	248,3436
11.	157,536	308,7991
12.	360,894	198,9661



Gambar 4. 31 Perbandingan Hasil Prediksi Model 3 Tanpa Normalisasi 2021

Tahapan selanjutnya mengevaluasi model menggunakan metrik evaluasi dengan menggunakan *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Hasil metrik evaluasi termuat dalam tabel 4.56.

Tabel 4. 56 Hasil Evaluasi Model 3 Tanpa Normalisasi 2021

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
3 Tanpa Normalisasi 2021	111,91196686907973	14988,532992548737	122.42766432693526

4.4.2.4 Tahun 2022

4.4.2.4.1 Dengan Normalisasi

Pada pengujian ini ditambahkan normalisasi data pada proses *preprocessing*. Terdapat tiga model yang akan diujikan. Perbedaan antar tiga model tersebut didasarkan pada perbandingan rasio *split data*. Setelah dilakukan

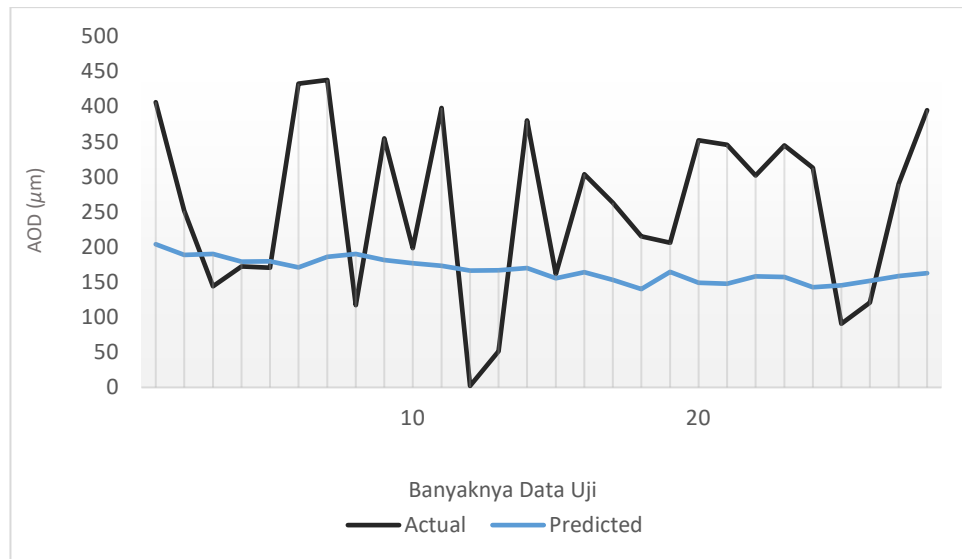
implementasi metode *linear regression* didapatkanlah hasil prediksi. Kemudian hasil prediksi tersebut dikembalikan ke skala semula untuk dilakukan evaluasi metrik dari hasil prediksi tersebut.

A. Model 1

Model 1 menggunakan *split data* dengan perbandingan 75% sebagai data *train* dan 25% sebagai data *test*. Berdasarkan dari split data tersebut, dari total 111 baris data, data yang di *training* sebanyak 83 baris data dan 28 baris data untuk di uji. Dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel 4.57 dan gambar 4.32. Di gambar 4.32, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna biru.

Tabel 4. 57 Perbandingan Hasil Prediksi Model 1 Normalisasi 2022

No.	<i>Actual</i> AOD (μm)	<i>Predicted</i> AOD (μm)
1.	406,109	203,9663
2.	251,639	188,8789
3.	144	190,3051
...
26.	121	151,7886
27.	289,818	158,6154
28.	395	162,7366



Gambar 4. 32 Perbandingan Hasil Prediksi Model 1 Normalisasi 2022

Selanjutnya, dilakukan evaluasi pada model. Hasil evaluasi metrik dari pengujian model 1 ditampilkan pada tabel 4.58. yang mencakup *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, dan *Root Mean Squared Error (RMSE)*.

Tabel 4. 58 Hasil Evaluasi Model 1 Normalisasi 2022

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
1 Normalisasi 2022	126,6416543421 2062	22463,20757434 735	149,87730840373 186

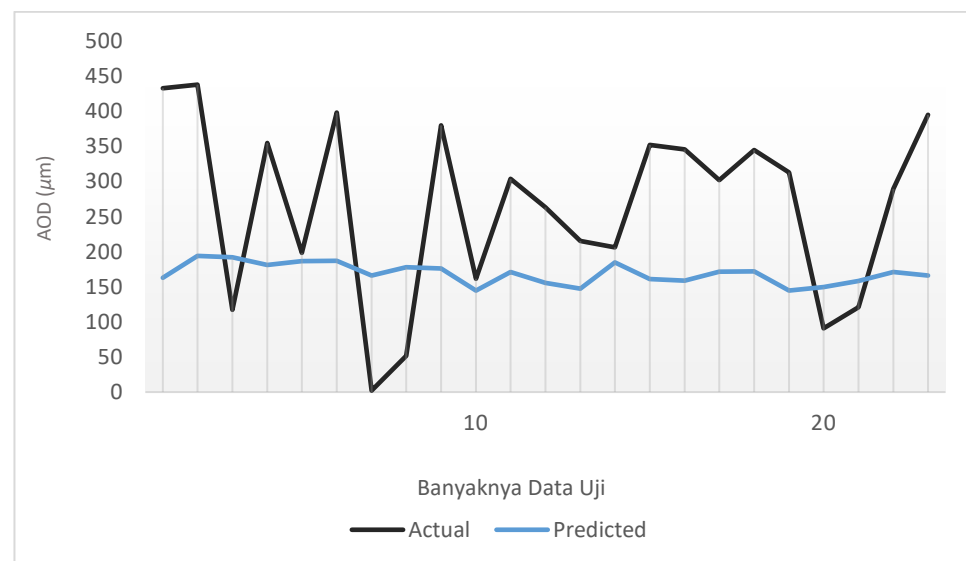
B. Model 2

Model 2 menggunakan *split data* menggunakan rasio 80% *train data* dan 20% *test data*. Dari total 111 baris data, 88 baris data *train* dan 23 baris data *test*. Selanjutnya, dilakukan proses prediksi pada pada model.

Dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel 4.59 dan gambar 4.33. Di gambar 4.33, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna biru.

Tabel 4. 59 Perbandingan Hasil Prediksi Model 2 Normalisasi 2022

No.	<i>Actual</i> AOD (μm)	<i>Predicted</i> AOD (μm)
1.	406,109	203,9663
2.	251,639	188,8789
3.	144	190,3051
...
21.	121	158,2786
22.	289,818	171,0374
23.	395	166,1674



Gambar 4. 33 Perbandingan Hasil Prediksi Model 2 Normalisasi 2022

Langkah berikutnya, model tersebut akan diuji menggunakan metrik evaluasi yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Hasil dari pengujian model ini ditampilkan pada tabel 4.60.

Tabel 4. 60 Hasil Evaluasi Model 2 Normalisasi 2022

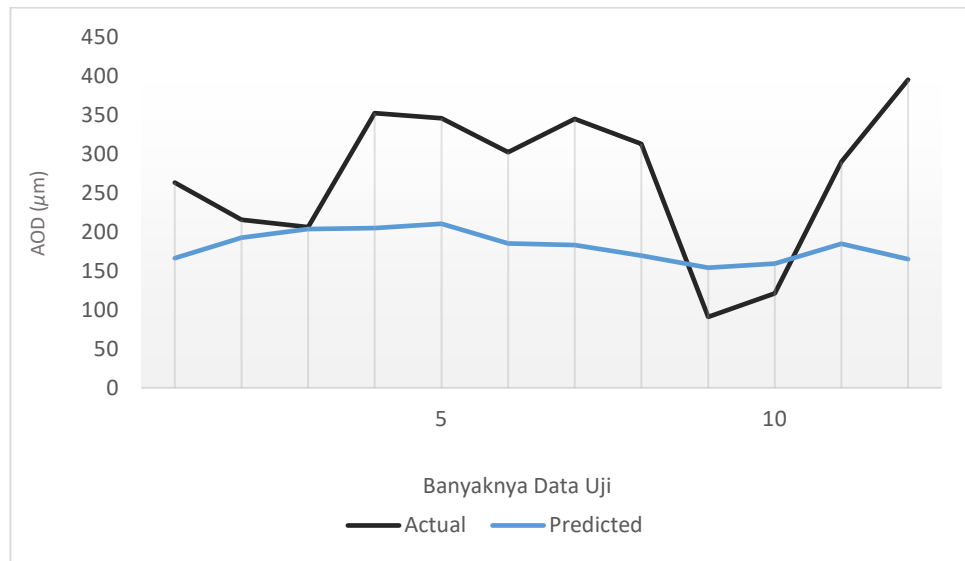
Model	Metrik Evaluasi		
	MAE	MSE	RMSE
2 Normalisasi 2022	135,5909932300 5236	23898,67144045 2457	154,5919514090 3182

C. Model 3

Pada model 3 *split data* diterapkan dengan rasio 90% *train data* dan 10% *test data*. Dari total 111 baris data, 99 baris data digunakan untuk proses *training model* dan 12 baris data untuk menguji model. Hasil dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel 4.61 dan gambar 4.34. Di gambar 4.34, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna biru.

Tabel 4. 61 Perbandingan Hasil Prediksi Model 3 Normalisasi 2022

No.	<i>Actual AOD</i> (μm)	<i>Predicted AOD</i> (μm)
1.	263,085	166,3037
2.	215,273	192,6601
3.	206	203,563
...
10.	121	159,1882
11.	289,818	184,4936
12.	395	164,8795



Gambar 4. 34 Perbandingan Hasil Prediksi Model 3 Normalisasi 2022

Tahapan selanjutnya mengevaluasi model menggunakan metrik evaluasi dengan menggunakan *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Hasil metrik evaluasi termuat dalam tabel 4.62.

Tabel 4. 62 Hasil Evaluasi Model 3 Normalisasi 2022

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
3 Normalisasi 2022	105,0503972287 3141	14224,96665052 9937	119,26846461043 228

4.4.2.4.2 Tanpa Normalisasi

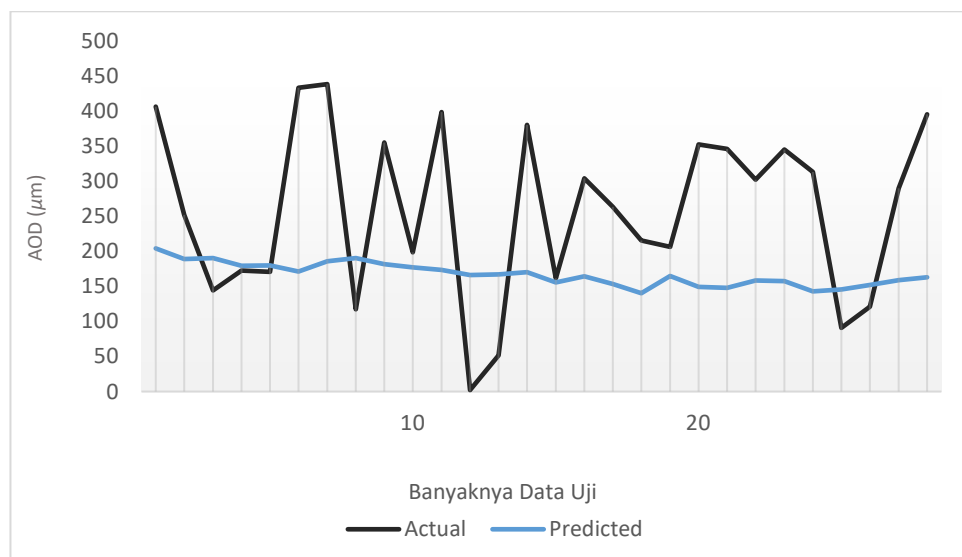
Pengujian dilakukan dengan melewati tahap normalisasi data. Pengujian ini juga dilakukan menggunakan tiga model uji. Ketiga model ini ditentukan berdasarkan *split data* yang dilakukan di tahap *preprocessing*.

A. Model 1

Model 1 menggunakan *split data* dengan perbandingan 75% *train data* dan 30% *test data*. Berdasarkan dari split data tersebut, dari total 111 baris data, data yang di *training* sebanyak 83 baris data dan 28 baris data untuk di uji. Dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel 4.63 dan gambar 4.35. Di gambar 4.35, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna biru.

Tabel 4. 63 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi 2022

No.	Actual AOD (μm)	Predicted AOD (μm)
1.	406,109	203,9663
2.	251,639	188,8789
3.	144	190,3051
...
26.	121	151,7886
27.	289,818	158,6154
28.	395	162,7366



Gambar 4. 35 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi 2022

Selanjutnya, dilakukan evaluasi pada model. Hasil evaluasi metrik dari pengujian model 1 ditampilkan pada tabel 4.64. yang mencakup *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE).

Tabel 4. 64 Hasil Evaluasi Model 1 Tanpa Normalisasi 2022

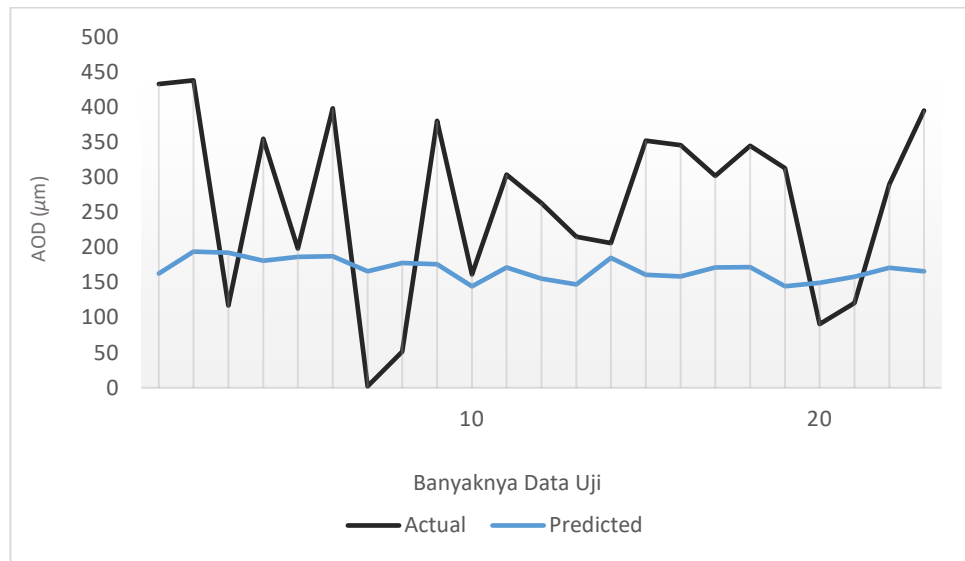
Model	Metrik Evaluasi		
	MAE	MSE	RMSE
1 Tanpa Normalisasi 2022	126,64165434212062	22463,207574347347	149,87730840373183

B. Model 2

Model 2 menerapkan rasio *split data* dengan 80% *train data* dan 20% *test data*. Dari total 111 baris data, 88 baris *train data* dan 23 baris *test data*. Selanjutnya, dilakukan proses prediksi pada pada model. Dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel 4.65 dan gambar 4.36. Di gambar 4.36, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna biru.

Tabel 4. 65 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi 2022

No.	<i>Actual AOD</i> (μm)	<i>Predicted AOD</i> (μm)
1.	433	162,9884
2.	438	194,0066
3.	117,103	192,2543
...
21.	121	158,2786
22.	289,818	171,0374
23.	395	166,1674



Gambar 4. 36 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi 2022

Langkah berikutnya, model tersebut akan diuji menggunakan metrik evaluasi yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Hasil dari pengujian model ini ditampilkan pada tabel 4.66

Tabel 4. 66 Hasil Evaluasi Model 2 Tanpa Normalisasi 2022

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
2 Tanpa Normalisasi 2022	135,5909932300 522	23898,6714404 524	154,59195140903 165

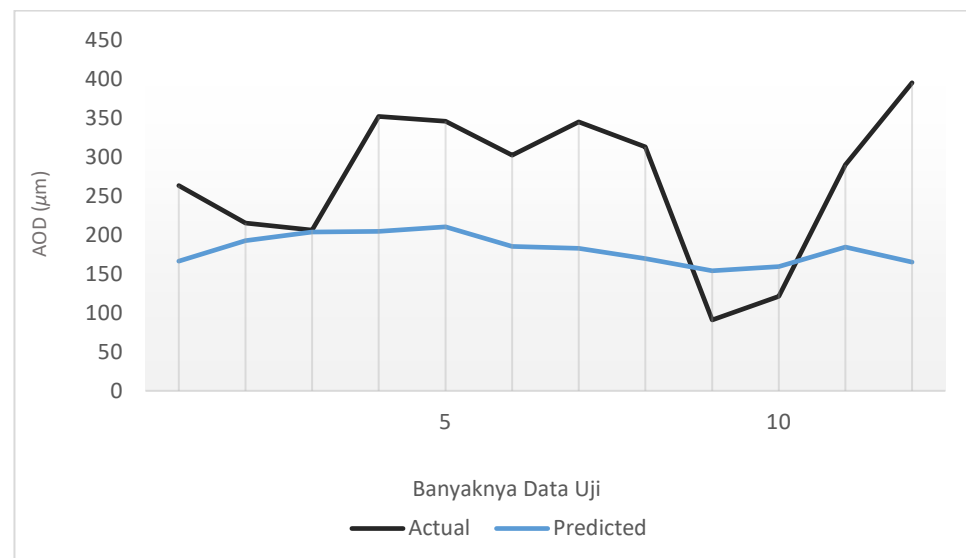
C. Model 3

Pada model 3 *split data* diterapkan dengan rasio 90% *train data* dan 10% *test data*. Dari total 111 baris data, 99 baris data digunakan pada proses *training model* dan 12 baris data digunakan pada *testing model*. Hasil dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel 4.67 dan gambar 4.37. Pada

gambar 4.37, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna biru.

Tabel 4. 67 Perbandingan Hasil Prediksi Model 3 Tanpa Normalisasi 2022

No.	Actual AOD (μm)	Predicted AOD (μm)
1.	263,085	166,3037
2.	215,273	192,6601
3.	206	203,563
...
10.	121	159,1882
11.	289,818	184,4936
12.	395	164,8795



Gambar 4. 37 Perbandingan Hasil Prediksi Model 3 Tanpa Normalisasi 2022

Tahapan selanjutnya mengevaluasi model menggunakan metrik evaluasi dengan menggunakan *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Hasil metrik evaluasi termuat dalam tabel 4.68.

Tabel 4. 68 Hasil Evaluasi Model 3 Tanpa Normalisasi 2022

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
3 Tanpa Normalisasi 2022	105,30795013142705	14993,001312140694	122,44591178206275

4.4.2.5 Tahun 2023

4.4.2.5.1 Dengan Normalisasi

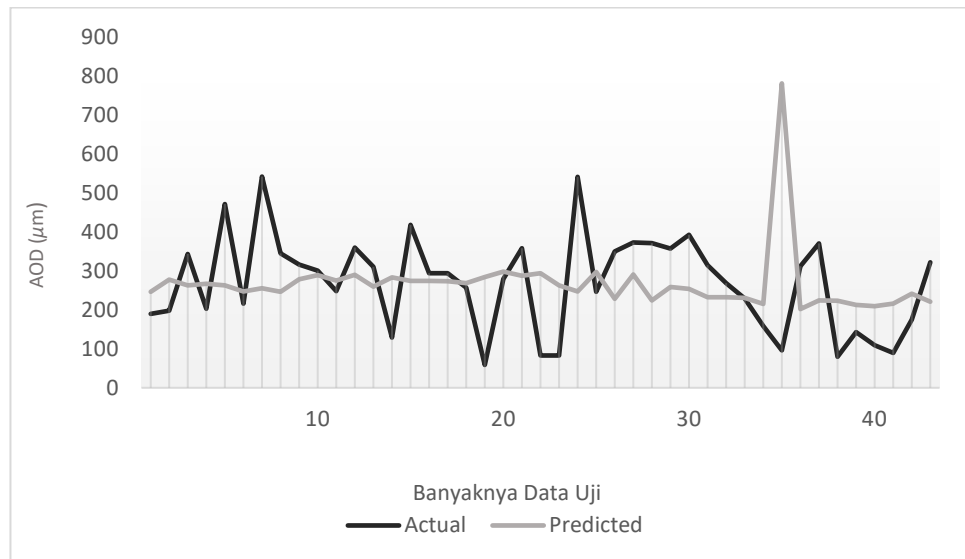
Pada pengujian ini ditambahkan normalisasi data pada proses *preprocessing*. Terdapat tiga model yang akan diujikan. Perbedaan antar tiga model tersebut didasarkan pada perbandingan rasio *split data*. Setelah dilakukan implementasi metode *linear regression* didapatkanlah hasil prediksi. Kemudian hasil prediksi tersebut dikembalikan ke skala semula untuk dilakukan evaluasi metrik dari hasil prediksi tersebut.

A. Model 1

Model 1 menggunakan *split data* dengan perbandingan 75% data *train* dan 25% data *test*. Berdasarkan dari split data tersebut, dari total 170 baris data, data yang di *training* sebanyak 127 baris data dan 43 baris data untuk di uji. Dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel 4.69 dan gambar 4.38. Di gambar 4.38, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna abu-abu.

Tabel 4. 69 Perbandingan Hasil Prediksi Model 1 Normalisasi 2023

No.	Actual AOD (μm)	Predicted AOD (μm)
1.	189,294	246,0763
2.	197,373	277,707
3.	343,409	262,5938
...
41.	89,437	215,8955
42.	174,716	241,143
43.	322	220,3506



Gambar 4. 38 Perbandingan Hasil Prediksi Model 1 Normalisasi 2023

Selanjutnya, dilakukan evaluasi pada model. Hasil evaluasi metrik dari pengujian model 1 ditampilkan pada tabel 4.70. yang mencakup *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, dan *Root Mean Squared Error (RMSE)*.

Tabel 4. 70 Hasil Evaluasi Model 1 Normalisasi 2023

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
1 Normalisasi 2023	112,0510305115 5108	25262,63619635 5757	158,9422416991 6492

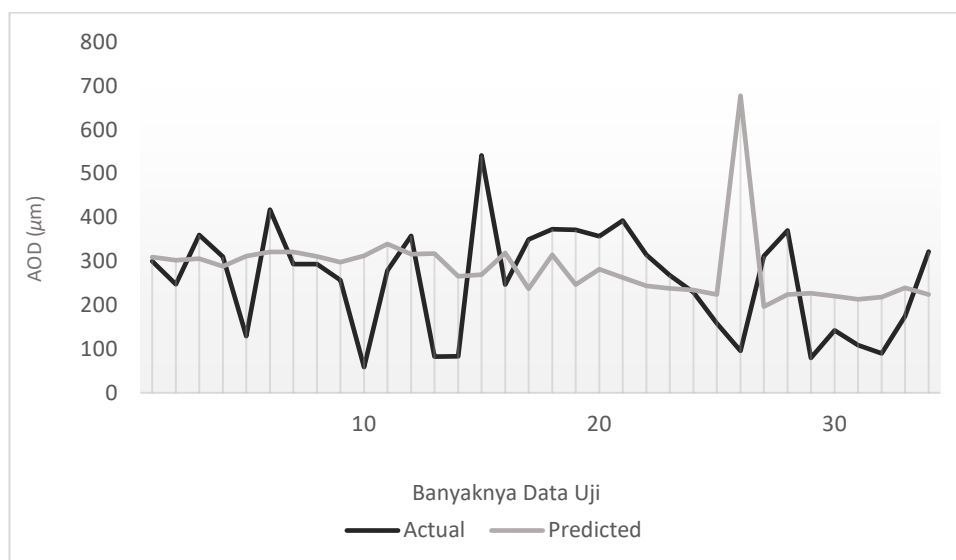
B. Model 2

Model 2 membagi data menggunakan rasio 80% *train data* dan 20% *test data*. Dari total 170 baris data, 136 baris data menjadi data *train* dan 34 baris data menjadi data *test*. Selanjutnya, dilakukan proses prediksi pada pada model. Dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel 4.71 dan gambar

4.39. Pada gambar 4.39, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna abu-abu.

Tabel 4. 71 Perbandingan Hasil Prediksi Model 2 Normalisasi 2023

No.	Actual AOD (μm)	Predicted AOD (μm)
1.	299,977	309,6828
2.	247,62	302,4451
3.	359,588	306,0672
...
32.	89,437	217,913
33.	174,716	239,2382
34.	322	223,9283



Gambar 4. 39 Perbandingan Hasil Prediksi Model 2 Normalisasi 2023

Langkah berikutnya, model tersebut akan diuji menggunakan metrik evaluasi yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Hasil dari pengujian model ini ditampilkan pada tabel 4.72.

Tabel 4. 72 Hasil Evaluasi Model 2 Normalisasi 2023

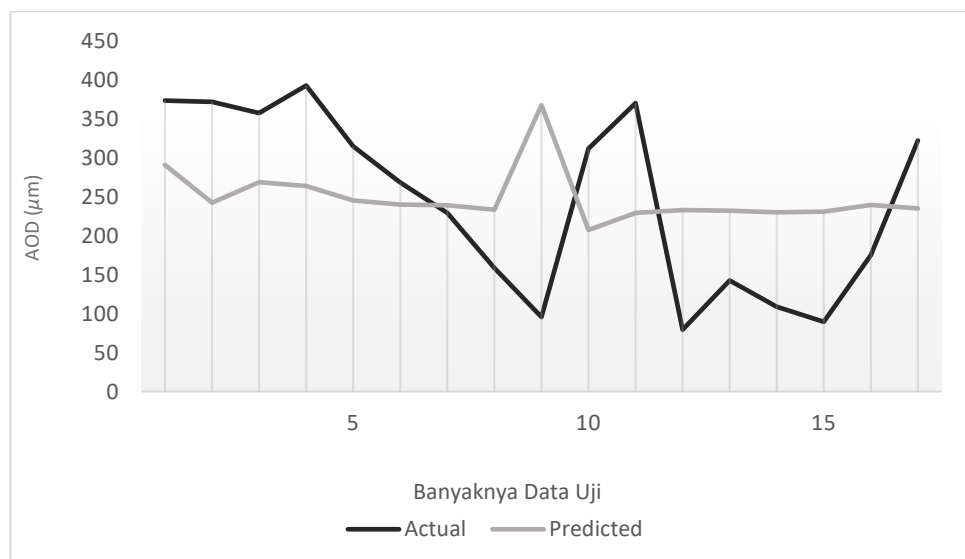
Model	Metrik Evaluasi		
	MAE	MSE	RMSE
2 Normalisasi 2023	110,5903874787 3225	23404,105339982 056	152,98400354279 548

C. Model 3

Pada model 3 *split data* diterapkan dengan rasio 90% *train data* dan 10% *test data*. Dari total 170 baris data, 153 baris data untuk proses *training model* dan 17 baris data untuk *testing model*. Hasil dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel 4.73 dan gambar 4.40. Di gambar 4.40, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna abu-abu.

Tabel 4. 73 Perbandingan Hasil Prediksi Model 3 Normalisasi 2023

No.	Actual AOD (μm)	Predicted AOD (μm)
1.	373	290,4927
2.	371,337	242,3197
3.	357,034	268,3492
...
15.	89,437	230,8
16.	174,716	239,2334
17.	322	234,9056



Gambar 4. 40 Perbandingan Hasil Prediksi Model 3 Normalisasi 2023

Tahapan selanjutnya mengevaluasi model menggunakan metrik evaluasi dengan menggunakan *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Hasil metrik evaluasi termuat dalam tabel 4.74.

Tabel 4. 74 Hasil Evaluasi Model 3 Normalisasi 2023

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
3 Normalisasi 2023	105,0503972287 3141	14224,96665052 9937	119,26846461043 228

4.4.2.5.2 Tanpa Normalisasi

Pengujian dilakukan dengan melewati tahap normalisasi data. Pengujian ini juga dilakukan menggunakan tiga model uji. Ketiga model ini ditentukan berdasarkan *split data* yang dilakukan di tahap *preprocessing*.

A. Model 1

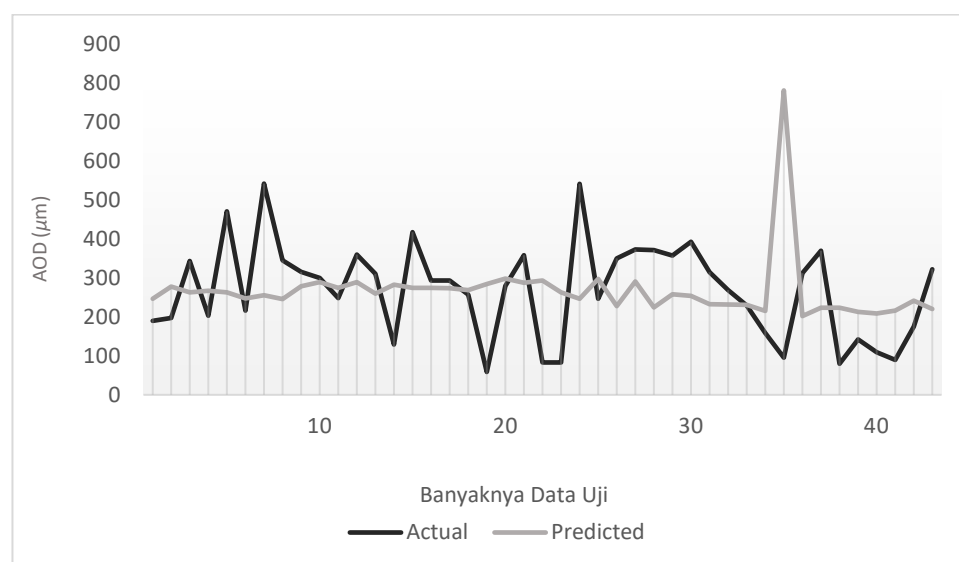
Model 1 menggunakan *split data* dengan perbandingan 75% *train data* dan 25% sebagai *test data*. Berdasarkan dari split data tersebut, dari total 170 baris data, data yang di *training* sebanyak 127 baris data dan 43 baris data untuk di uji. Dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel 4.75 dan gambar 4.41. Di gambar 4.41, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna abu-abu.

Selanjutnya, dilakukan evaluasi pada model. Hasil evaluasi metrik dari pengujian model 1 ditampilkan pada tabel 4.76. yang mencakup *Mean*

Absolute Error (MAE), Mean Squared Error (MSE), dan Root Mean Squared Error (RMSE).

Tabel 4. 75 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi 2023

No.	Actual AOD (μm)	Predicted AOD (μm)
1.	189,294	246,0763
2.	197,373	277,707
3.	343,409	262,5938
...
41.	89,437	215,8955
42.	174,716	241,143
43.	322	220,3506



Gambar 4. 41 Perbandingan Hasil Prediksi Model 1 Tanpa Normalisasi 2023

Tabel 4. 76 Hasil Evaluasi Model 1 Tanpa Normalisasi 2023

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
1 Tanpa Normalisasi 2023	112,0510305115 5133	25262,6361963 5603	158,94224169916 578

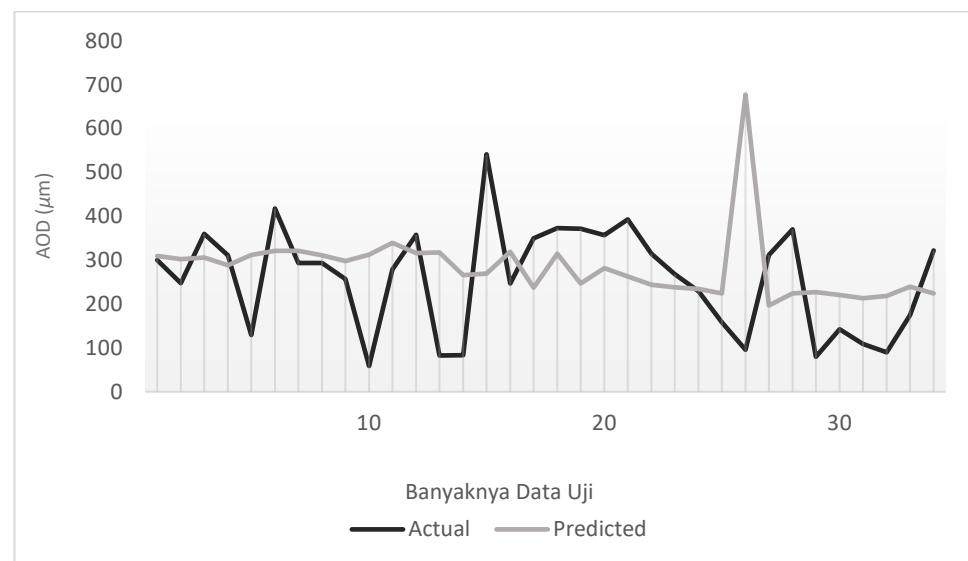
B. Model 2

Model 2 membagi data menggunakan rasio 80% *train data* dan 20% *data test data*. Dari total 170 baris data, 136 baris data menjadi data

train dan 34 baris data menjadi data *test*. Selanjutnya, dilakukan proses prediksi pada pada model. Dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel 4.77 dan gambar 4.42. Di gambar 4.42, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna abu-abu.

Tabel 4. 77 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi 2023

No.	Actual AOD (μm)	Predicted AOD (μm)
1.	299,977	309,6828
2.	247,62	302,4451
3.	359,588	306,0672
...
32.	89,437	217,913
33.	174,716	239,2382
34.	322	223,9283



Gambar 4. 42 Perbandingan Hasil Prediksi Model 2 Tanpa Normalisasi 2023

Langkah berikutnya, model tersebut akan diuji menggunakan metrik evaluasi yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Hasil dari pengujian model ini ditampilkan pada tabel 4.78.

Tabel 4. 78 Hasil Evaluasi Model 2 Tanpa Normalisasi 2023

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
2 Tanpa Normalisasi 2022	110,590387478 73205	23404,1053399 81874	152,984003542 79488

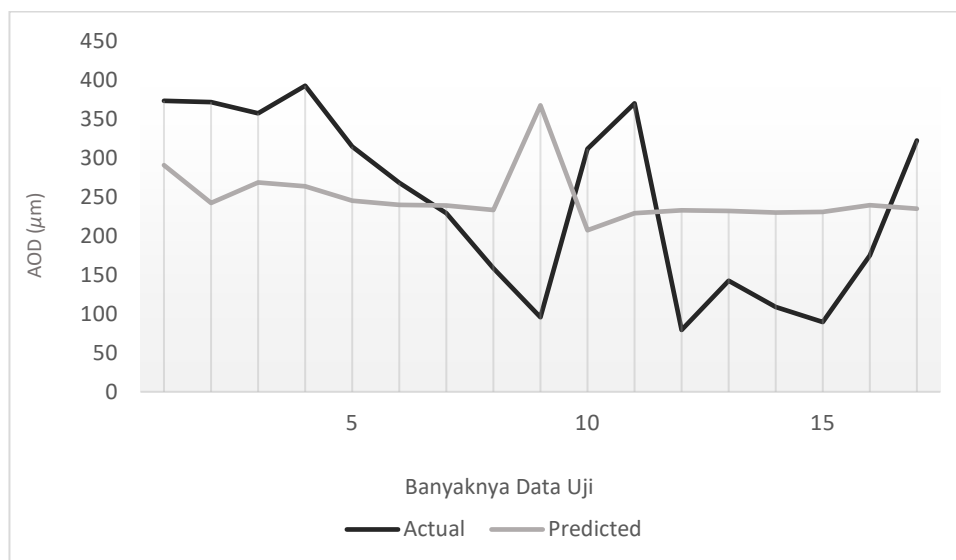
C. Model 3

Pada model 3 *split data* diterapkan dengan rasio 90% *train data* dan 10% *test data*. Dari total 170 baris data, 153 baris data untuk proses *training model* dan 17 baris data untuk *testing model*. Hasil dari pengujian model ini didapatkan hasil prediksi yang dapat dibandingkan dengan nilai sebenarnya di tabel 4.79 dan gambar 4.43. Di gambar 4.43, nilai sebenarnya berupa garis berwarna hitam dan hasil prediksi berwarna abu-abu.

Tahapan selanjutnya mengevaluasi model menggunakan metrik evaluasi dengan menggunakan *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan *Root Mean Squared Error* (RMSE). Hasil metrik evaluasi termuat dalam tabel 4.80

Tabel 4. 79 Perbandingan Hasil Prediksi Model 3 Tanpa Normalisasi 2023

No.	<i>Actual AOD</i> (μm)	<i>Predicted AOD</i> (μm)
1.	373	290,4927
2.	371,337	242,3197
3.	357,034	268,3492
...
15.	89,437	230,8
16.	174,716	239,2334
17.	322	234,9056



Gambar 4. 43 Perbandingan Hasil Prediksi Model 3 Tanpa Normalisasi 2023

Tabel 4. 80 Hasil Evaluasi Model 3 Tanpa Normalisasi 2023

Model	Metrik Evaluasi		
	MAE	MSE	RMSE
3 Tanpa Normalisasi 2022	105,05039722873	14224,9666505299	119,26846461043
	146	63	239

4.5 Evaluasi

Subbab ini akan membahas tentang hasil uji coba yang telah dilakukan pada *subbab* sebelumnya. Pembahasan hasil uji coba akan dibahas sesuai dengan skenario pengujian dan didasarkan jenis matrik evaluasi secara berurutan dimulai dari MAE, MSE, lalu RMSE. Hal ini bertujuan untuk mempermudah dalam menentukan model yang memiliki performa terbaik.

4.5.1 Mean Absolut Error (MAE)

Mean Absolut Error (MAE) mengukur kesalahan prediksi dengan menghitung selisih absolut dari hasil prediksi dengan nilai sebenarnya lalu mengambil rata-rata dari hasil selisih tersebut. Semakin kecil nilai MAE suatu

model berarti semakin baik model tersebut untuk melakukan prediksi (Nurani et al., 2023).

Nilai MAE data lima tahun (2019-2023) yang didapatkan dari pengujian semua model yang telah dilakukan pada subbab sebelumnya disajikan menjadi satu tabel yaitu, tabel 4.81. Hasil MAE pada semua model memiliki selisih yang sangat kecil. Pada model 1 antara kedua skenario memiliki selisih hanya sebesar 0,000000000000001, dengan model normalisasi lebih rendah. Pada model 2 antara kedua skenario memiliki selisih hanya sebesar 0,000000000000002, dengan model normalisasi lebih rendah. Sedangkan pada model 3 antara kedua skenario memiliki selisih hanya sebesar 0,000000000000002, dengan model tanpa normalisasi lebih rendah. Diantara ketiga model, model 2 memiliki nilai MAE terendah. Dari kedua skenario model 2, model dengan normalisasi memiliki nilai yang lebih rendah. Jadi, model terbaik berdasarkan kategori matrik evaluasi *Mean Absolute Error* (MAE) pada data lima tahun adalah model 2 dengan normalisasi.

Tabel 4. 81 MAE Lima Tahun

Mean Absolute Error (MAE)		
Model	Normalisasi	Tanpa Normalisasi
Model 1 (75/25)	86,82527641136865	86,82527641136866
Model 2 (80/20)	83,83957904259945	83,83957904259947
Model 3 (90/10)	93,99779032292813	93,99779032292811

Nilai MAE data pertahun yang didapatkan dari pengujian semua model yang telah dilakukan pada subbab sebelumnya disajikan menjadi satu tabel yaitu, tabel 4.82. Hasil MAE pada semua model di tiap tahun memiliki selisih yang sangat kecil antara menggunakan normalisasi ataupun tanpa normalisasi. Pada tahun 2019 nilai MAE terendah didapatkan saat menggunakan model 1 dengan

normalisasi, yaitu sebesar 96,47619133208842. Nilai MAE terendah tahun 2019 ditunjukkan warna oranye. Pada tahun 2020 nilai MAE terendah didapatkan saat menggunakan model 2 dengan normalisasi, yaitu sebesar 78,84943592611091. Nilai MAE terendah tahun 2020 ditunjukkan warna kuning. Pada tahun 2021 nilai MAE terendah didapatkan saat menggunakan model 1 tanpa normalisasi yaitu sebesar, 92,94774237090463. Nilai MAE terendah tahun 2021 ditunjukkan warna hijau. Pada tahun 2022 nilai MAE terendah didapatkan saat menggunakan model 3 tanpa normalisasi, yaitu sebesar 105,30795013142705. Nilai MAE terendah tahun 2022 ditunjukkan warna biru. Pada tahun 2023 nilai MAE terendah didapatkan saat menggunakan model 3 dengan normalisasi, yaitu sebesar 105,05039722873141. Nilai MAE terendah tahun 2023 ditunjukkan warna abu-abu.

Tabel 4. 82 MAE Pertahun

Mean Absolute Error (MAE) Pertahun			
Model	Tahun	Normalisasi	Tanpa Normalisasi
Model 1 (75/25)	2019	96,47619133208842	96,47619133208848
	2020	86,08453440384996	86,08453440385001
	2021	92,94774237090468	92,94774237090463
	2022	126,64165434212062	126,64165434212062
	2023	112,05103051155108	112,05103051155133
Model 2 (80/10)	2019	109,34379176846039	109,34379176846011
	2020	78,84943592611091	78,84943592611101
	2021	102,04726270077727	102,04726270077735
	2022	135,59099323005236	135,5909932300522
	2023	110,59038747873225	110,59038747873205
Model 3 (90/10)	2019	101,085667425938	101,08566742593803
	2020	85,90267303245933	85,90267303245948
	2021	111,91196686907978	111,91196686907973
	2022	105,30795013142712	105,30795013142705
	2023	105,05039722873141	105,05039722873146

4.5.2 Mean Squared Error (MSE)

Mean Squared Error (MSE) mengukur kinerja model dengan menghitung kuadrat selisih dari hasil prediksi dengan hasil sebenarnya. Model dengan hasil *Mean Squared Error* (MSE) yang paling rendah adalah yang terbaik.

Nilai MSE data lima tahun (2019-2023) yang didapatkan dari pengujian semua model yang telah dilakukan pada subbab sebelumnya disajikan menjadi satu tabel yaitu, tabel 4.83. Hasil MSE pada model 1 dan 3 memiliki selisih yang sangat kecil antar skenarionya, sedangkan model 2 memiliki hasil yang sama. Pada model 1 antara kedua skenario memiliki selisih hanya sebesar 0,000000000000003, dengan model normalisasi lebih rendah. Pada model 3 antara kedua skenario memiliki selisih hanya sebesar 0,000000000000002, dengan model tanpa normalisasi lebih rendah. Dari ketiga model tersebut, model 2 dengan split data 80/20 memiliki nilai MSE terendah yaitu sebesar 11064,628777469545. Sehingga model terbaik pada data lima tahun berdasarkan MSE adalah model 2 dengan atau tanpa normalisasi data.

Tabel 4. 83 MSE Data Lima Tahun

Mean Squared Error (MSE)		
	Normalisasi	Tanpa Normalisasi
Model 1 (75/25)	11417,131300788868	11417,131300788871
Model 2 (80/20)	11140,332314620859	11140,332314620859
Model 3 (90/10)	12929,743349831315	12929,743349831317

Nilai MSE data pertahun yang didapatkan dari pengujian semua model yang telah dilakukan pada subbab sebelumnya disajikan menjadi satu tabel yaitu, tabel 4.84. Hasil MSE pada semua model di tiap tahun memiliki selisih yang sangat kecil antara menggunakan normalisasi atau tanpa normalisasi. Pada tahun

2019 nilai MSE terendah didapatkan saat menggunakan model 1 dengan normalisasi, yaitu sebesar 13977,482765012692. Nilai MSE terendah tahun 2019 ditunjukkan warna oranye. Pada tahun 2020 nilai MSE terendah didapatkan saat menggunakan model 2 dengan normalisasi, yaitu sebesar 7934,003612790981. Nilai MSE terendah tahun 2020 ditunjukkan warna kuning. Pada tahun 2021 nilai MSE terendah didapatkan saat menggunakan model 1 tanpa normalisasi, yaitu sebesar 13952,856468910024. Nilai MSE terendah tahun 2021 ditunjukkan warna hijau. Pada tahun 2022 nilai MSE terendah didapatkan saat menggunakan model 3 tanpa normalisasi, yaitu sebesar 14993,001312140694. Nilai MSE terendah tahun 2022 ditunjukkan warna biru. Pada tahun 2023 nilai MSE terendah didapatkan saat menggunakan model 3 dengan normalisasi, yaitu sebesar 14224,966650529937. Nilai MSE terendah tahun 2023 ditunjukkan warna abu-abu.

Tabel 4. 84 MSE Data Pertahun

Mean Squared Error (MSE) Pertahun			
Model	Tahun	Normalisasi	Tanpa Normalisasi
Model 1 (75/25)	2019	13977,482765012692	13977,4827650127
	2020	10020,486182971323	10020,486182971335
	2021	13952,85646891004	13952,856468910024
	2022	22463,20757434735	22463,207574347347
	2023	25262,636196355757	25262,63619635603
Model 2 (80/10)	2019	17363,30202679657	17363,302026796486
	2020	7934,003612790981	7934,003612791004
	2021	16213,707665872265	16213,707665872287
	2022	23898,671440452457	23898,6714404524
	2023	23404,105339982056	23404,105339981874
Model 3 (90/10)	2019	14082,781306841593	14082,7813068416
	2020	9220,032129764912	9220,032129764943
	2021	14988,532992548746	14988,532992548737
	2022	14993,001312140708	14993,001312140694
	2023	14224,966650529937	14224,966650529963

4.5.3 Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) mengukur kinerja model dengan menggunakan akar kuadrat dari *Mean Squared Error* (MSE). Model yang memiliki hasil prediksi yang baik adalah model dengan nilai RMSE terendah.

Nilai MAE data lima tahun (2019-2023) yang didapatkan dari pengujian semua model yang telah dilakukan pada subbab sebelumnya disajikan menjadi satu tabel yaitu, tabel 4.85. Ketiga model memiliki hasil RMSE yang sama pada kedua skenarionya. Model 2 memiliki hasil RMSE sebesar 105,5477726653711. Sehingga, model terbaik pada data lima tahun berdasarkan evaluasi RMSE adalah model 2 dengan atau tanpa normalisasi data.

Tabel 4. 85 RMSE Data Lima Tahun

Root Mean Squared Error (RMSE)		
	Normalisasi	Tanpa Normalisasi
Model 1 (75/25)	106,85097706988395	106,85097706988395
Model 2 (80/20)	105,5477726653711	105,5477726653711
Model 3 (90/10)	113,70902932411003	113,70902932411003

Nilai RMSE data pertahun yang didapatkan dari pengujian semua model yang telah dilakukan pada subbab sebelumnya disajikan menjadi satu tabel yaitu, tabel 4.86. Hasil RMSE pada semua model di tiap tahun memiliki selisih yang sangat kecil antara menggunakan normalisasi ataupun tanpa normalisasi. Pada tahun 2019 nilai RMSE terendah didapatkan saat menggunakan model 1 dengan normalisasi, yaitu sebesar 118,22640468614738. Nilai RMSE terendah tahun 2019 ditunjukkan warna oranye. Pada tahun 2020 nilai RMSE terendah didapatkan saat menggunakan model 2 dengan atau tanpa normalisasi, yaitu sebesar 89,0730240465148. Nilai RMSE terendah tahun 2020 ditunjukkan warna

kuning. Pada tahun 2021 nilai RMSE terendah didapatkan saat menggunakan model 1 tanpa normalisasi, yaitu sebesar 118,12220988836106. Nilai RMSE terendah tahun 2021 ditunjukkan warna hijau. Pada tahun 2022 nilai RMSE terendah didapatkan saat menggunakan model 3 dengan atau tanpa normalisasi, yaitu sebesar 122,44591178206281. Nilai RMSE terendah tahun 2022 ditunjukkan warna biru. Pada tahun 2023 nilai RMSE terendah didapatkan saat menggunakan model 3 dengan atau tanpa normalisasi, yaitu sebesar 119,26846461043228. Nilai RMSE terendah tahun 2023 ditunjukkan warna abu-abu.

Tabel 4. 86 RMSE Data Pertahun

Root Mean Square Error (RMSE) Pertahun			
Model	Tahun	Normalisasi	Tanpa Normalisasi
Model 1 (75/25)	2019	118,22640468614738	118,22640468614742
	2020	100,10237850806205	100,10237850806212
	2021	118,12220988836113	118,12220988836106
	2022	149,87730840373186	149,87730840373183
	2023	158,94224169916492	158,94224169916578
Model 2 (80/10)	2019	131,76988285187358	131,76988285187358
	2020	89,0730240465148	89,0730240465148
	2021	127,3330580245063	127,3330580245063
	2022	154,59195140903165	154,59195140903165
	2023	152,98400354279488	152,98400354279488
Model 3 (90/10)	2019	118,6708949441336	118,6708949441336
	2020	96,02099837933842	96,02099837933842
	2021	122,4276643269353	122,4276643269353
	2022	122,44591178206281	122,44591178206281
	2023	119,26846461043228	119,26846461043228

4.6 Integrasi Islam

4.6.1 Muamalah Mu'Allah

Seiring dengan perubahan zaman, manusia disibukkan dengan banyak aktivitas mulai dari yang berskala kecil hingga besar. Tanpa disadari, aktivitas

manusia ini secara langsung atau tidak langsung memberi dampak pada lingkungan. Seperti halnya dengan udara, banyak dari kita mungkin tidak terlalu menyadari tentang perubahan udara yang kita hirup dalam beberapa tahun terakhir. Allah berfirman dalam Al-Qur'an surah An-Nahl ayat 48.

أَوَلَمْ يَرَوْا إِلَىٰ مَا خَلَقَ اللَّهُ مِن شَيْءٍ يَتَتَمَيُّزُ ظِلُّهُ عَنِ الْيَمِينِ وَالشَّمَالِ سُجَّدًا لِلَّهِ وَهُمْ دُخْرُونَ

“Apakah mereka tidak memperhatikan segala sesuatu yang diciptakan Allah, bayang-bayangnya berbolak-balik ke kanan dan ke kiri (dalam keadaan) sujud kepada Allah, sedangkan mereka rendah hati.” (Q.S An-Nahl 16:48).

Menurut tafsir dari Kementrian Agama RI, Apakah mereka tidak memperhatikan benda yang diciptakan Allah mulai dari yang besar maupun kecil, dari bergerak maupun diam, hidup maupun mati, yang bayang-bayangnya berbolak-balik ke kanan dan ke kiri, semua dalam keadaan sujud dan tunduk kepada Allah dan bersikap rendah hati menerima ketetapan itu. Tidakkah mereka memperhatikan sehingga mereka sadar dan mau beriman kepada nabi Muhammad dan tidakkah mereka memperhatikan pula bahwa segala apa yang ada di langit dan di bumi hanya bersujud kepada Allah' demikian juga para malaikat yang tidak pernah menyombongkan diri dan membangkang perintah Allah.

Dari tafsir diatas dapat dijelaskan pentingnya kita menyadari segala sesuatu yang ada disekitar kita dan beriman kepada Allah dan Rosul-Nya. Seperti halnya menyadari perubahan pada udara. Udara yang kita hirup sudah berubah kandungan yang ada di dalamnya. Sebagai contoh yang mudah dikenali adalah warna hitam pada udara yang berada di jalan yang ramai kendaraan berlalu-lalang. Udara tersebut akan terasa menyesakkan jika dihirup terus menerus. Dengan

menyadari perubahan tersebut diharapkan manusia dapat menanggulangi dampaknya, mengurangi faktor-faktor penyebabnya, tidak menyalahkan ketetapan Allah serta tetap beriman dan bertaqwa kepada Allah dan Rosul-nya.

4.6.2 Muamalah Mu'Annas

Upaya mewujudkan kualitas udara yang baik dapat dimulai dari diri sendiri. Selanjutnya dapat mengajak orang lain untuk bersama-sama menjaga kualitas udara, seperti menghindari perbuatan buruk yang dapat memperburuk kualitas udara saat ini. Sehingga, kita mendapatkan kebaikan dan keselamatan hidup. Seperti firman Allah dalam Quran Surah Al-Imran ayat ke 104:

وَلْتَكُنْ مِنْكُمْ أُمَّةٌ يَدْعُونَ إِلَى الْخَيْرِ وَيَأْمُرُونَ بِالْمَعْرُوفِ وَيَنْهَوْنَ عَنِ الْمُنْكَرِ ؕ وَأُولَٰئِكَ هُمُ الْمُفْلِحُونَ

“Dan hendaklah ada di antara kamu segolongan umat yang menyeru kepada kebajikan, menyuruh kepada yang ma'ruf dan mencegah dari yang munkar; merekalah orang-orang yang beruntung.” (Q.S Al-Imran 3:104).

Menurut tafsir dari Kementrian Agama RI, ayat ini memerintahkan orang mukmin untuk mengajak manusia kepada kebaikan, menyuruh perbuatan makruf, dan mencegah perbuatan mungkar. Menyerukan kebaikan-kebaikan yaitu petunjuk-petunjuk Allah dan mencegah dari yang mungkar, yaitu sesuatu yang dipandang buruk dan diingkari oleh akal sehat. Sungguh mereka yang menjalankan hal tersebut mempunyai kedudukan tinggi di hadapan Allah dan mereka itulah orang-orang yang beruntung karena mendapatkan keselamatan di dunia dan akhirat.

4.6.3 Muamalah Mu'Alam

Penyebab penurunan kualitas udara bisa disebabkan karena fenomena alam seperti gunung meletus. Akan tetapi faktor dominan dari penurunan kualitas udara disebabkan oleh manusia. Tindakan manusia yang seperti industrialisasi dan penggunaan bahan bakar fosil membuat polutan-polutan pencemar udara menjadi meningkat. Peningkatan ini tentunya akan memberikan dampak buruk dari penurunan kualitas udara seperti perubahan iklim yang dapat memusnahkan spesies hewan dan tumbuhan. Oleh sebab itu, kita harus menjaga kelestarian bumi. Seperti firman Allah dalam surat Al-Qasas ayat 77:

وَأَبْتَعْ فِي مِمَّا آتَاكَ اللَّهُ الدَّارَ الْآخِرَةَ ۖ وَلَا تَنْسَ نَصِيبَكَ مِنَ الدُّنْيَا ۖ وَأَحْسِنَ كَمَا أَحْسَنَ اللَّهُ إِلَيْكَ ۖ وَلَا تَتَّبِعِ الْفُسَادَ فِي الْأَرْضِ ۖ إِنَّ اللَّهَ لَا يُحِبُّ الْمُفْسِدِينَ

“Dan carilah pada apa yang telah dianugerahkan Allah kepadamu (kebahagiaan negeri akhirat, dan janganlah kamu melupakan bahagianmu dari (kenikmatan) duniawi dan berbuat baiklah (kepada orang lain) sebagaimana Allah telah berbuat baik, kepadamu, dan janganlah kamu berbuat kerusakan di (muka) bumi. Sesungguhnya Allah tidak menyukai orang-orang yang berbuat kerusakan.” (Q.S. AL-Qasas 28:77).

Tasir al Muyassar menjelaskan bahwa carilah pahala akhirat dengan harta benda yang Allah berikan dan jangan lupakan bagianmu untuk bersenang-senang didunia dengan hal yang halal dan tidak berlebihan. Berbuat baiklah dengan memberikan sedekah seperti Allah yang memberimu harta yang banyak. Jangan mencari hal yang diharamkan Allah dengan melakukan tindakan berbuat kerusakan dimuka bumi dan menganiaya kaum. Sesungguhnya Allah tidak

menyukai orang-orang yang berbuat kerusakan dan Dia akan membalas mereka atas amal perbuatan buruk mereka.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, didapatkan bahwa hasil MAE, MSE, RMSE antara menggunakan data yang dinormalisasi atau data tanpa normalisasi memiliki hasil yang relatif sama pada tiap model. Hal ini berarti ada atau tidaknya normalisasi hanya memiliki pengaruh yang sangat kecil pada hasil evaluasi.

Dari hasil evaluasi dengan data lima tahun (2019-2023), model terbaik didapatkan pada model 2 dengan rasio pembagian data 80% data *training* dan 20% data *testing*. Model 2 memiliki hasil evaluasi MAE, MSE, dan RMSE terendah dibandingkan dengan model lainnya. Nilai MAE sebesar 83,83957904259947, MSE sebesar 11140,332314620859, dan RMSE sebesar 105,5477726653711. Sedangkan pada hasil evaluasi dengan data tahunan menunjukkan bahwa model tahun 2020 memiliki hasil terendah MAE, MSE, dan RMSE dibandingkan tahun lainnya. Model terbaik tahun 2020 didapatkan dari model 2 dengan normalisasi dengan rasio pembagian data 80% data *training* dan 20% data *testing*. Model 2 memiliki hasil evaluasi MAE sebesar 78,84943592611091, MSE sebesar 7934,003612790981, dan RMSE sebesar 89,0730240465148.

Berdasarkan hasil evaluasi keseluruhan model, hanya model tahun 2020 yang memiliki hasil evaluasi MAE, MSE, dan RMSE lebih rendah dari hasil evaluasi model data lima tahun. Sehingga pengujian yang sesuai untuk penelitian ini dengan menggunakan data lima tahunan. Selain itu, penggunaan metode linear

regression masih kurang dalam melakukan prediksi, hasil prediksi memiliki selisih yang cukup jauh dari nilai aktual sehingga membuat nilai MAE, MSE, dan RMSE menjadi tinggi.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, terdapat beberapa saran yang dapat digunakan sebagai berikut:

1. Penambahan faktor pengaruh PM.2.5 lainnya seperti parameter-parameter meteorologi (suhu, arah mata angin, dan kelembapan).
2. Perubahan lokasi penelitian seperti lokasi yang memiliki perbedaan aktivitas penduduk yang berpengaruh pada kualitas udara.
3. Pengambilan langkah preventif agar kualitas udara tidak menurun.

DAFTAR PUSTAKA

- Alita, D., Putra, A. D., & Darwis, D. (2021). Analysis of classic assumption test and multiple linear regression coefficient test for employee structural office recommendation. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(3), 295. <https://doi.org/10.22146/ijccs.65586>
- Amansyah, I., Indra, J., & Juwita, A. R. (2024). Prediksi Penjualan Kendaraan Menggunakan Regresi Linier: Studi Kasus pada Industri Otomotif di Indonesia. *INNOVATIVE: Journal Of Sosial Science Research*, 4(4), 1199–1216. <https://j-innovative.org/index.php/Innovative>
- Anggraini, T. S., Irie, H., Sakti, A. D., & Wikantika, K. (2024). Machine learning-based global air quality index development using remote sensing and ground-based stations. *Environmental Advances*, 15. <https://doi.org/10.1016/j.envadv.2023.100456>
- Arisandi, D., Salamun, S., & Putra, A. R. (2023). Prediksi Penerimaan Siswa Baru dengan Metode Single Exponential Smoothing Melalui Metrik Evaluasi MAD, MSE dan MAPE. *Journal of Information System Research (JOSH)*, 4(4), 1197–1204. <https://doi.org/10.47065/josh.v4i4.3658>
- Astriyani, M., Laela, I. N., Lestari, D. P., Anggraeni, L., & Astuti, T. (2022). Analisis Klasifikasi Data Kualitas Udara DKI Jakarta Menggunakan Algoritma C.45. *Jurnal Sistem & Teknologi Komunikasi*, 6(1), 36–41.
- Ayuningtyas, N., & Yustanti, W. (2024). Semi-Supervised Learning pada Pelabelan dalam Klasifikasi Multi-Label Data Teks. *Journal of Informatics and Computer Science*, 06(1), 240–248.
- Azmi, B. N., Hermawan, A., & Avianto, D. (2023). Analisis Pengaruh Komposisi Data Training dan Data Testing pada Penggunaan PCA dan Algoritma Decision Tree untuk Klasifikasi Penderita Penyakit Liver. *JTIM: Jurnal Teknologi Informasi Dan Multimedia*, 4(4), 281–290. <https://doi.org/10.35746/jtim.v4i4.298>
- Baskara, A. K., Nazir, A., Irsyad, M., Yusra, & Isnani, F. (2023). Implementasi Data Mining Memprediksi Penjualan Crude Palm Oil Berdasarkan Kapasitas Tangki Menggunakan Multiple Linear Regression. *Jurnal Sistem Komputer Dan Informatika*, 4(3), 493–502. <https://doi.org/10.30865/json.v4i3.5665>
- Dian, F., Emilia, R., Gery, G. P., & Padjadjaran, U. (2023). *Klasifikasi Tingkat Pencemaran Udara Kota Jakarta Tahun 2021 Menggunakan Algoritma Decission Tree*. <https://www.data.jakarta.go.id/>
- Fikri, A. A., Darmawan, A., Hilmanto, R., Banuwa, I. S., Agustiono, A., & Agustiana, L. (2022). Pemanfaatan platform Google Earth Engine dalam Pemantauan Perubahan Tutupan Lahan di Taman Hutan Raya Wan Abdul

- Rachman. *Journal of Forest Science Avicennia*, 5(1), 46–57. <https://doi.org/10.22219/avicennia.v5i1.19938>
- Garajeh, M. K., Laneve, G., Rezaei, H., Sadeghnejad, M., Mohamadzadeh, N., & Salmani, B. (2023). Monitoring Trends of CO, NO₂, SO₂, and O₃ Pollutants Using Time-Series Sentinel-5 Images Based on Google Earth Engine. *Pollutants*, 3(2), 255–279. <https://doi.org/10.3390/pollutants3020019>
- Ginting, L. M., Sigiro, M. M., Manurung, E. D., & Sinurat, J. J. P. (2021). Perbandingan Metode Algoritma Support Vector Regression dan Multiple Linear Regression Untuk Memprediksi Stok Obat. *Journal of Applied Technology and Informatics*, 1(2), 29–34. <http://journal-jati.del.ac.id/>
- Greenstone, M., & Fan, Q. (2019). *Kualitas Udara Indonesia yang Memburuk dan Dampaknya terhadap Harapan Hidup*.
- Hikmah, N., Bahri, S., & Irwansyah, I. (2022). Jaringan Syaraf Tiruan untuk Memprediksi Kadar Polutan Ozon di Kota Mataram. *Eigen Mathematics Journal*, 5(2), 70–85. <https://doi.org/10.29303/emj.v5i2.129>
- Husna, S., Sakdiah, Idayati, R., Safri, M., Mulia, V. D., & Bakhtiar. (2022). Hubungan Sumber Polutan dalam Rumah dengan Kejadian Infeksi Saluran Pernafasan Akut pada Balita di Wilayah Kerja Puskesmas Lubuk Sanai, Mukomulyo, Bengkulu. *Jurnal Kedokteran Nangroe Medika*, 5(1).
- IQAir. (2022). *PM2.5*. <https://www.iqair.com/Id/Newsroom/Pm2-5>. <https://www.iqair.com/id/newsroom/pm2-5>
- Khan, S. I., & Hoque, A. S. M. L. (2020). SICE: an improved missing data imputation technique. *Journal of Big Data*, 7(37), 1–21. <https://doi.org/10.1186/s40537-020-00313-w>
- Kumar, A., & Jamatia, A. (2023). Prediction of Air Quality Using Machine Learning. *Smart Innovation, Systems and Technologies*, 326 *SIST*, 199–209. https://doi.org/10.1007/978-981-19-7513-4_18
- Latief, M. A., & Karyanti, Y. (2022). Data Mining & Analytic Forecasting Indeks Standar Pencemar Udara Jakarta Menggunakan Metode Linear Regression (Studi Kasus: Dataset Indeks Standar Pencemar Udara Jakarta 2021). *JOSR: Journal of Social Research*, 1(10), 1164–1176. <http://https://ijsr.internationaljournallabs.com/index.php/ijsrhttp://ijsr.internationaljournallabs.com/index.php/ijsr>
- Lewar, T. T., & Darmawan, S. (2022). Estimasi Aerosol Optical Depth (AOD) Secara Time Series Menggunakan Citra Satelit Modis (Studi Kasus: Pulau Jawa). In *FTSP Series*.
- Lumeno, F., Patras, Li., & Lisi, F. (2016). Pengukuran Indeks Polusi Pada Sistem Minahasa Berdasarkan Nilai Esdd Dan Nsd. *E-Journal Teknik Elektro Dan Komputer*, 5(2), 50–58.

- Maherdyta, N. R., Syafitri, A., Septywantoro, F., Kejora, P. A., Gulo, S. D., & Sulistiyorini, D. (2022). Analisis Risiko Kesehatan Lingkungan Paparan Gas Nitrogen Dioksida (NO₂) dan Sulfur Dioksida (SO₂) pada Masyarakat di Wilayah Yogyakarta. *Jurnal Sanitasi Lingkungan*, 2(1), 51–59. <https://doi.org/10.36086/jsl.v2i1.1040>
- Maimaiti, A. (2023). How Does Human Thinking Activity Impact the Health and Stability of The Atmosphere? *International Journal of Education and Humanities*, 9(1), 51–65.
- Martha, A. (2022). *Direktori Perusahaan Industri Besar dan Sedang Kota Malang 2022* (R. C. Novita, Ed.; 1st ed.). BPS Kota Malang.
- Masacgi, G. N., & Rohman, M. S. (2023). Optimasi Model Algoritma Klasifikasi menggunakan Metode Bagging pada Stunting Balita. *Edumatic: Jurnal Pendidikan Informatika*, 7(2), 455–464. <https://doi.org/10.29408/edumatic.v7i2.23812>
- Masito, A. (2018). Analisis Resiko Kualitas Udara Ambien (NO₂ dan SO₂) dan Gangguan Pernapasan Pada Masyarakat di Wilayah Kalianak Surabaya. *Jurnal Kesehatan Lingkungan*, 10(4), 394–401.
- Miskhahuddin. (2019). Manusia dan Lingkungan Hidup Dalam Al-Qur'an. *Al-Mu'ashirah*, 16(2), 210–227.
- Nur, N., Rosidin, R., & Dirgawati, M. (2023). Kajian Literatur : Kemampuan Aerosol Optical Depth (AOD) dalam Memprediksi Kualitas Udara Parameter PM_{2,5}. *FTSP Series*, 1975–1983.
- Nurani, A. T., Setiawan, A., & Susanto, B. (2023). Perbandingan Kinerja Regresi Decision Tree dan Regresi Linear Berganda untuk Prediksi BMI pada Dataset Asthma. *Jurnal Sains Dan Edukasi Sains*, 6(1), 34–43. <https://doi.org/10.24246/juses.v6i1p34-43>
- Nurramadhani, A., Riandi, R., Permanasari, A., & Suwama, I. R. (2024). Low-Carbon Food Consumption for Solving Climate Change Mitigation: Literature Review with Bibliometric and Simple Calculation Application for Cultivating Sustainability Consciousness in facing Sustainable Development Goals (SDGs). *Indonesian Journal of Science and Technology*, 9(2), 261–268. <https://doi.org/10.17509/ijost.v9i1.67302>
- Oktafiani, R., Hermawan, A., & Avianto, D. (2023). Pengaruh Komposisi Split data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning. *Jurnal Sains Dan Informatika*. <https://doi.org/10.34128/jsi.v9i1.622>
- Pebralia, J. (2022). Analisis Curah Hujan Menggunakan Machine Learning Metode Regresi Linier Berganda Berbasis Python dan Jupyter Notebook Rainfall Analysis using Machine Learning-Multiple Linear Regression Method Based

- on Python and Jupyter Notebook. *Jurnal Ilmu Fisika Dan Pembelajarannya*, 6(2), 23–30. <http://jurnal.radenfatah.ac.id/index.php/jifp/>
- Peraturan Badan Meteorologi, Klimatologi, Dan Geofisika Republik Indonesia Nomor 2 Tahun 2020 Tentang Penyediaan Dan Penyebaran Informasi Kualitas Udara, Pub. L. No. 2, 1 (2020).
- Peraturan Menteri Lingkungan Hidup Dan Kehutanan Republik Indonesia Nomor P.14/MENLHK/SETJEN/KUM.1/7/2020 Tentang Indeks Standar Pencemar Udara, Pub. L. No. 14, 1 (2020).
- Plevris, V., Solorzano, G., Bakas, N. P., Amine, M. El, & Seghier, M. E. A. B. S. (2022). *Investigation of Performance Metrics in Regression Analysis and Machine Learning-Based Prediction Models*. 1–25.
- Prasetya, M. R. A., Priyatno, A. M., & Nurhaeni. (2023). Penanganan Imputasi Missing Values pada Data Time Series dengan Menggunakan Metode Data Mining. *Jurnal Informasi Dan Teknologi*, 5(2), 56–62. <https://doi.org/10.37034/jidt.v5i1.324>
- Rambing, V. V, Umboh, J. M. L., & Warouw, F. (2022). Literature Review: Gambaran Risiko Kesehatan pada Masyarakat akibat Paparan Gas Karbon Monoksida (CO). *Jurnal KESMAS*, 11(4), 95–102.
- Ramdani, F. (2021). *Google Earth Engine: Metode, Teknik, dan Aplikasi* (1st ed.). UB Press.
- Safira, M. D., Syafiuddin, A., Fasya, A. H. Z., & Setianto, B. (2022). Kualitas Udara di Kawasan Industri di Berbagai Lokasi di Indonesia. *Jurnal Public Health*, 9(2), 38–47.
- Sukoco, B. (2022). *Kajian Pemanfaatan Teknologi Google Earth Engine Untuk Bidang Penginderaan Jauh*. Universitas Lampung.
- Vermaysha, A., & Nurmalitasari. (2023). *Prediksi Harga Rumah di Kabupaten Karanganyar Menggunakan Metode Regresi Linear Sistem Informasi*.
- Wibowo, R. A., & Kurniawan, A. A. (2020). Analisis Korelasi Dalam Penentuan Arah Antar Faktor Pada Pelayanan Angkutan Umum di Kota Magelang. *Theta Omega: Journal of Electrical Engineering*, 1–6.
- Yusuf, R., & Huda, A. A. (2023). Deteksi Emosi Wajah Menggunakan Metode Backpropagation Detection of Facial Emotions Using The Backpropagation Method. In *JACIS: Journal Automation Computer Information System* (Vol. 3, Issue 2).
- Zhao, Q., Yu, L., Li, X., Peng, D., Zhang, Y., & Gong, P. (2021). Progress and trends in the application of google earth and google earth engine. *Remote Sensing*, 13(18), 1–21. <https://doi.org/10.3390/rs13183778>