



**IMPLEMENTASI ALGORITMA *MULTIPLE LINEAR REGRESSION*
UNTUK MEMPREDIKSI TEMPERATUR UDARA BERDASARKAN
KADAR ZAT POLUTAN DI KOTA TANGERANG SELATAN**

SKRIPSI

**TEDJA DIAH RANI OCTAVIA
NIM. 2010511130**

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA
2024**



**IMPLEMENTASI ALGORITMA *MULTIPLE LINEAR REGRESSION*
UNTUK MEMPREDIKSI TEMPERATUR UDARA BERDASARKAN
KADAR ZAT POLUTAN DI KOTA TANGERANG SELATAN**

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana
Komputer**

**TEDJA DIAH RANI OCTAVIA
NIM. 2010511130**

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA
2024**

PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya saya sendiri dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Tedja Diah Rani Octavia

NIM : 2010511130

Tanggal : 29 Juli 2024

Judul : Implementasi Algoritma Multiple Linear Regression untuk
Memprediksi Temperatur Udara Berdasarkan Kadar Zat Polutan
di Kota Tangerang Selatan

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Depok, 29 Juli 2024



Tedja Diah Rani Octavia

**PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK
KEPENTINGAN AKADEMIS**

Sebagai civitas akademika Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan di bawah ini:

Nama : Tedja Diah Rani Octavia

NIM : 2010511130

Fakultas : Ilmu Komputer

Program Studi : S1 Informatika

Demi Pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

**Implementasi Algoritma Multiple Linear Regression untuk Memprediksi
Temperatur Udata Berdasarkan Kadar Zat Polutan di Kota Tangerang
Selatan**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini, Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Tugas Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Depok, 29 Juli 2024



Tedja Diah Rani Octavia

LEMBAR PENGESAHAN

Tugas akhir ini diajukan oleh:

Nama : Tedja Diah Rani Octavia

NIM : 2010511130


Program Studi : S1 Informatika

Judul Skripsi : Implementasi Algoritma *Multiple Linear Regressoin* Untuk Memprediksi
Temperatur Udara Berdasarkan Kadar Zat Polutan di Kota Tangerang
Selatan

Telah berhasil dipertahankan di hadapan tim penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.


Bayu Hananto, S.Kom., M.Kom.

Penguji 1


Neny Rosmawati, S.Kom., M.Kom.

Pembimbing 1



Prof. Dr. Ir. Supriyanto, S.T., M.Sc., IPM
Dekan


Muhammad Panji Muslim, S.Pd., M.Kom.

Penguji 2


Ati Zaidiah, S.Kom., MTI

Pembimbing 2


Dr. Widya Cholil, M.K.T.
Kepala Program Studi

Ditetapkan di : Jakarta

Tanggal Persetujuan : 12 Juli 2024

IMPLEMENTASI ALGORITMA *MULTIPLE LINEAR REGRESSION* UNTUK MEMPREDIKSI TEMPERATUR UDARA BERDASARKAN KADAR ZAT POLUTAN DI KOTA TANGERANG SELATAN

TEDJA DIAH RANI OCTAVIA

ABSTRAK

Pemanasan global menjadi salah satu masalah paling besar yang dihadapi oleh manusia saat ini. Fenomena ini telah berdampak luas pada berbagai aspek, terutama aspek lingkungan. Secara global, dampak yang akan timbul akibat dari pemanasan global adalah meningkatnya suhu bumi. Pada April 2023, sebagian besar negara di Asia Selatan, termasuk Bangladesh, Myanmar, India, China, Thailand, dan Laos, mengalami gelombang panas dengan suhu melebihi 40°C. Di Indonesia, suhu maksimum harian tercatat mencapai 37,2°C di stasiun pengamatan BMKG di Ciputat, dengan beberapa lokasi mencatat suhu antara 34°C hingga 36°C. Pemanasan global dan perubahan iklim disebabkan oleh meningkatnya konsentrasi gas rumah kaca terutama CO₂, CH₄, N₂O, CFC, dan Ozon (O₃) pada atmosfer bumi. Pada penelitian ini diusulkan model prediksi rata-rata temperatur udara dengan memperhatikan pengaruh kadar zat polutan, seperti CO, NO, NO₂, O₃, SO₂, PM_{2.5}, PM₁₀, dan NH₃. Prediksi temperatur udara pada penelitian ini menggunakan dataset iklim harian dan kadar zat polutan di Kota Tangerang Selatan yang diambil pada rentang waktu 25 November 2020 hingga 31 Desember 2023. Pemodelan prediksi pada penelitian ini menggunakan algoritma *Multiple Linear Regression* dengan pembagian data 80% data latih dan 20% data uji. Dari 8 uji coba, model keenam merupakan model terbaik dengan nilai k=8 dan fitur *RH_avg*, *RR*, *ss*, *ddd_car*, *ff_avg*, *no2*, *o3*, dan *pm10*. Hasil evaluasi model keenam diperoleh nilai R² sebesar 0.72749, MAE sebesar 0.55593, dan MSE sebesar 0.50078 dengan persentase error atau MAPE sebesar 1,99806%.

Kata Kunci: Temperatur Udara, Zat Polutan, *Multiple Linear Regression*

IMPLEMENTATION OF MULTIPLE LINEAR REGRESSION ALGORITHM TO PREDICT AIR TEMPERATURE BASED ON POLLUTANT LEVELS IN SOUTH TANGERANG CITY

TEDJA DIAH RANI OCTAVIA

ABSTRACT

Global warming is one of the biggest problems faced by humans today. This phenomenon has had a wide impact on various aspects, especially the environment. In April 2023, most countries in South Asia, including Bangladesh, Myanmar, India, China, Thailand, and Laos, experienced a heat wave with temperatures exceeding 40°C. In Indonesia, the daily maximum temperature was recorded at 37.2°C at the BMKG observation station in Ciputat, with some locations recording temperatures between 34°C and 36°C. Global warming and climate change are caused by the increasing concentration of greenhouse gases, especially CO₂, CH₄, N₂O, CFCs, and Ozone (O₃) in the earth's atmosphere. In this study, a prediction model of average air temperature is proposed by considering the influence of pollutant levels, such as CO, NO, NO₂, O₃, SO₂, PM_{2.5}, PM₁₀, and NH₃. The prediction of air temperature in this study uses daily climate datasets and pollutant levels in South Tangerang City taken from November 25, 2020 to December 31, 2023. Prediction modeling in this study uses the Multiple Linear Regression algorithm with 80% training data and 20% test data. From 8 trials, the sixth model is the best model with a value of $k = 8$ and features RH_avg, RR, ss, ddd_car, ff_avg, no2, o3, and pm10. The sixth model evaluation results obtained an R² value of 0.72749, MAE of 0.55593, and MSE of 0.50078 with a percentage error or MAPE of 1,99806%.

Keywords: Air Temperature, Pollutants, Multiple Linear Regression

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan atas kehadiran Allah SWT karena berkat kehendak dan izin-Nya, penulis dapat menyelesaikan skripsi ini yang berjudul “Implementasi Algoritma Multiple Linear Regression untuk Memprediksi Temperatur Udara Berdasarkan Kadar Zat Polutan di Kota Tangerang Selatan” dengan baik dan tepat waktu. Dalam penyusunan skripsi ini, penulis mendapat banyak dukungan baik materi maupun moral dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih sebesar-besarnya terutama kepada:

1. Ayah dan Ibu penulis yang selalu merawat, mendidik, mendukung dan mendoakan penulis.
2. Ibu Neny Rosmawarni, S.Kom, M.Kom., Ibu Ati Zaidiah S.Kom., MTI., dan Almarhumah Ibu Yuni Widiastiwi S.Kom., M.Si. selaku dosen pembimbing yang telah membimbing, mengarahkan, dan memberikan saran serta masukan dalam penulisan skripsi ini.
3. Sahabat penulis, Adelle dan Dimas yang selalu ada di sisi penulis serta memberikan dukungan mental dan moral.
4. Teman-teman kuliah seperjuangan penulis, Amel, Alysha, Ica, Refa, Savina, dan Yaasintha yang selama ini selalu menemani penulis selama masa perkuliahan.

Penulis menyadari, skripsi ini masih belum sempurna dan memiliki kesalahan, baik dalam segi penyusunan, bahasa, hingga penulisan. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun agar dapat menjadi acuan bagi penulis untuk menjadi lebih baik lagi di masa mendatang. Penulis berharap skripsi ini dapat memberikan manfaat dan dapat dijadikan media pembelajaran bagi para pembaca.

Depok, 20 April 2024

Penulis

DAFTAR ISI

PERNYATAAN ORISINALITAS	iii
PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS	iv
LEMBAR PENGESAHAN	v
ABSTRAK	vi
ABSTRACT	vii
KATA PENGANTAR	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN.....	xiv
BAB I PENDAHULUAN.....	15
1.1 Latar Belakang	15
1.2 Rumusan Masalah	16
1.3 Tujuan Penelitian.....	16
1.4 Manfaat Penelitian.....	16
1.5 Batasan Masalah.....	17
BAB II TINJAUAN PUSTAKA.....	18
2.1 Kajian Literatur	18
2.2 Landasan Teori	24
2.2.1 Regresi Linear	24
2.2.2 <i>Multiple Linear Regression</i>	25
2.2.3 <i>Ordinary Least Squares</i>	26
2.2.4 Prediksi.....	27

2.2.5	<i>Exploratory Data Analysis</i>	27
2.2.6	Koefisien Determinasi.....	29
2.2.7	<i>Mean Absolute Error</i> dan <i>Mean Squared Error</i>	29
2.2.8	<i>Mean Absolute Percentage Error</i>	30
2.2.9	Pemanasan Global.....	31
2.2.10	Polutan	32
2.2.11	Temperatur	32
2.2.12	OpenWeather <i>Air Pollution</i> API	33
BAB III METODE PENELITIAN.....		35
3.1	Kerangka Berpikir	35
3.2	Identifikasi Masalah	35
3.3	Studi Literatur.....	36
3.4	Pengumpulan Data	36
3.5	Praproses Data	36
3.5.1	<i>Data Cleaning</i>	37
3.5.2	<i>Exploratory Data Analysis</i>	37
3.5.3	Transformasi Variabel.....	37
3.5.4	Pemilihan Variabel.....	38
3.5.5	Pembagian Data	38
3.6	Perancangan Model	39
3.7	Evaluasi Model.....	39
3.8	Perangkat Penelitian	39
3.8.1	Perangkat Keras	40
3.8.2	Perangkat Lunak.....	40
3.9	Jadwal Penelitian	40
BAB IV HASIL DAN PEMBAHASAN		42

4.1	Data	42
4.1.1	Data Iklim Harian.....	42
4.1.2	Data Kadar Zat Polutan.....	43
4.1.3	<i>Merging Data</i>	44
4.2	Exploratory Data Analysis	46
4.2.1	Heatmap	46
4.2.2	Scatter Plot	47
4.2.3	Histogram.....	47
4.3	Praproses Data	49
4.3.1	Drop Kolom	49
4.3.2	Imputasi <i>Missing Value</i>	49
4.3.3	Penanganan <i>Outliers</i>	50
4.3.4	Pembagian Data	51
4.3.5	Standardisasi	51
4.4	Pemodelan dan Evaluasi.....	55
4.4.1	Pemilihan Model Melalui Evaluasi Percobaan	55
4.4.2	Implementasi <i>Multiple Linear Regression</i>	68
4.4.3	Perhitungan Model Manual.....	71
4.5	Pengujian Data	85
4.6	Kompleksitas Model.....	89
BAB V KESIMPULAN		90
5.1	Kesimpulan.....	90
5.2	Saran.....	91
DAFTAR PUSTAKA		93
LAMPIRAN		97

DAFTAR TABEL

Tabel 2.1 Penelitian Relevan Terdahulu.....	18
Tabel 2.2 Skala <i>Air Quality Index</i> OpenWeather	33
Tabel 3.1 Jadwal Penelitian.....	40
Tabel 4.1 Data Iklim Harian.....	42
Tabel 4.2 Data Kadar Zat Polutan	44
Tabel 4.3 Data Gabungan.....	45
Tabel 4.4 Imputasi <i>Missing Value</i>	50
Tabel 4.5 Persentase Pembagian Data Latih dan Data Uji.....	51
Tabel 4.6 Data Latih yang Telah Distandardisasi.....	53
Tabel 4.7 Data Uji yang Telah Distandardisasi	54
Tabel 4.8 Hasil Evaluasi Semua Percobaan	66
Tabel 4.9 Hasil Perkalian Antar Variabel	74
Tabel 4.10 Hasil Penjumlahan Perkalian Antar Variabel	74
Tabel 4.11 Sampel Data Uji Penghitungan Manual	85
Tabel 4.12 Perbandingan Hasil Prediksi dengan Nilai Sesungguhnya	87

DAFTAR GAMBAR

Gambar 3.1. Kerangka Berpikir	35
Gambar 4.1. <i>Heatmap/Correlation Map</i>	46
Gambar 4.2. <i>Scatter Plot</i>	47
Gambar 4.3. Histogram.....	48
Gambar 4.4. Hasil Evaluasi Percobaan 1	55
Gambar 4.5. Hasil Evaluasi Percobaan 2	56
Gambar 4.6. Hasil Evaluasi Percobaan 3	58
Gambar 4.7. Fitur Percobaan 4	59
Gambar 4.8. Hasil Evaluasi Percobaan 4	59
Gambar 4.9. Fitur Percobaan 5	60
Gambar 4.10. Hasil Evaluasi Percobaan 5	60
Gambar 4.11. Fitur Percobaan 6.....	62
Gambar 4.12. Hasil Evaluasi Percobaan 6	62
Gambar 4.13. Fitur Percobaan 7	63
Gambar 4.14. Hasil Evaluasi Percobaan 7	63
Gambar 4.15. Fitur Percobaan 8	64
Gambar 4.16. Hasil Evaluasi Percobaan 8	65
Gambar 4.17. <i>Intercept</i> dan Koefisien Model.....	71
Gambar 4.18. Temperatur Pengujian Ke-1.....	86
Gambar 4.19. Temperatur Pengujian Ke-2.....	86
Gambar 4.20. <i>Mean Absolute Percentage Error</i> dengan Python.....	87
Gambar 4.21. Grafik Perbandingan Hasil Prediksi dengan Nilai Sesungguhnya ..	88
Gambar 4.22. Kompleksitas Waktu dan Memori Model	89

DAFTAR LAMPIRAN

Lampiran 1. Turnitin	97
Lampiran 2. Dataset	98
Lampiran 3. Kode Program.....	99