**PREDIKSI DAN ANALISIS POLA PERUBAHAN IKLIM MENGGUNAKAN ALGORITMA *GRADIENT BOOSTING***

**Pingki Ans Saputra1\*, Sandra Septi Irawan2, Rahmaddeni3, Robi Prianto4, Taupik Hidayat5**

1,2,3,4,5Teknik Informatika, Universitas Sains dan Teknologi Indonesia; Jl. Purwodadi Indah No.KM. 10, Pekanbaru, Riau

|  |  |
| --- | --- |
|  |  |
| Received: xxxx-xx-xx Accepted: xx-xx-xx  **Keywords:** analisis;  perubahan iklim;  Gradient Boosting; RMSE.  **Corespondent Email:** rahmaddeni@usti.ac.id | **Abstrak.** *Peningkatan suhu global merupakan indikator utama perubahan iklim yang memengaruhi pola cuaca, kenaikan permukaan laut, dan frekuensi kejadian cuaca ekstrem. Penelitian ini bertujuan memprediksi dan menganalisis pola perubahan iklim, khususnya suhu global, menggunakan algoritma Gradient Boosting. Dataset yang digunakan dalam penelitian ini diambil dari tahun 2020 sampai dengan 2024 dari platform Kaggle berjumlah 53 baris. Bedasarkan pengujian yang dilakukan dengan menggunakan Gradient Boosting hasil prediksi dengan menggunakan metrik evaluasi RMSE dan MAE menunjukkan bahwa metode ini menghasilkan error yang yang kecil pada CO2 Concentration (ppm), dengan RMSE nya 10.43, MAE nya adalah 8.92. Untuk Sea Surface Temp (°C), RMSE nya 8.57, MAE nya adalah 7.42, menunjukkan kesalahan prediksi suhu permukaan laut yang relatif kecil.* |
| **Abstract.** *Global temperature increase is a major indicator of climate change that affects weather patterns, sea level rise, and the frequency of extreme weather events. This study aims to predict and analyze climate change patterns, especially global temperature, using the Gradient Boosting algorithm. The dataset used in this study was taken from 2020 to 2024 from the Kaggle platform totaling 53 rows. Based on testing carried out using Gradient Boosting, the prediction results using the RMSE and MAE evaluation metrics show that this method produces a small error in CO2 Concentration (ppm), with an RMSE of 10.43, MAE of 8.92. For Sea Surface Temp (°C), the RMSE is 8.57, MAE is 7.42, indicating a relatively small sea surface temperature prediction error.* |
|  |  |

# PENDAHULUAN

Perubahan iklim global telah menjadi perhatian utama dalam beberapa dekade terakhir, karena dampaknya yang luas terhadap lingkungan, ekonomi, dan kesehatan manusia [1]. Salah satu aspek yang paling mencolok dari perubahan iklim adalah peningkatan suhu rata-rata global yang memengaruhi berbagai faktor seperti pola cuaca, kenaikan permukaan laut, dan kejadian cuaca ekstrem [2]. Dalam upaya memahami dinamika perubahan iklim dan merancang strategi mitigasi yang efektif, penting untuk mengembangkan model prediksi yang akurat mengenai suhu global yang dipengaruhi oleh berbagai variabel atmosfer.

Seiring dengan kemajuan teknologi, metode pembelajaran mesin semakin banyak digunakan dalam prediksi perubahan iklim, termasuk algoritma yang berbasis pada *ensemble learning* seperti *Gradient Boosting Machine* (GBM). *Gradient Boost Regresi* (GBR) adalah sejumlah keputusan untuk membuat pohon untuk klasifikasi dan regresi. Cara umum untuk bekerja dari dorongan gradien adalah dengan membuat urutan pohon di mana setiap pohon berfokus pada prediksi yang tersisa dari pohon sebelumnya*.* [3]. Selain itu, *Gradient Boosting* merupakan algoritma regresi yang fleksibel karena mengadopsi sifat-sifat dari .*regression* .*tree* yang memiliki interpretabilitas .tinggi, konsepnya yang sederhana dan efisiensi komputasi dengan pendekatan *boosting* yang dapat meningkatkan akurasi faktor mode [4]. Untuk menilai kinerja model, penelitian ini menggunakan. *mean square eror* (MSE) dan *root mean square error* (RMSE), yaitu perhitungan kesalahan dalam model regresi. Ini diberikan total waktu untuk kesalahan prediksi individu dan prediksi untuk membagi data [5].

Pada penelitian sebelumnya [6], yang berjudul “Prediksi Harga Mobil Menggunakan Algoritma Regressi dengan *Hyper*-Parameter *Tuning*” menggunakan *Gradient Boost Regression* yang memiliki nilai akurasi model sebesar 97% (setelah *tuning*) untuk melakukan prediksi. Pada penelitian selanjutnya [7], yang berjudul “Analisis Perbandingan Metode Regresi Linier, *Random Forest Regression* dan *Gradient Boosted Trees Regression Method* untuk Prediksi Harga Rumah” menunjukkan nilai RMSE *Gradient Boosted Trees Regression* sebesar 0,508 dengan nilai akurasi 75%. Penelitian yang dilakukan oleh [8], yang berjudul “Prediksi Faktor Penyebab Emisi CO2 pada Kendaraan Menggunakan *Gradient Boosting Regression*” mendapatkan nilai akurasi *Gradient boosting Regression* sebesar 99,43%. Penelitian lainnya dilakukan oleh [9], yang berjudul “Penerapan Algoritma *Gradient Boosting* pada Sinyal EEG sebagai Pengendali Kursi Roda” mendapatkan akurasi pelatihan sebesar 93.37% dan pengujian 93.33%. Berdasarkan permasalahan tersebut maka penelitian selanjutnya akan mencoba menerapkan algoritma *Gradient Boosting* untuk memprediksi dan menganalisis pola perubahan iklim.

Hasil yang diharapkan dari penelitian ini adalah model prediksi yang mampu mengidentifikasi faktor-faktor utama yang mempengaruhi perubahan suhu serta memberikan wawasan yang dapat digunakan untuk mendukung upaya mitigasi perubahan iklim global.

# TINJAUAN PUSTAKA

## Analisis

Analisis adalah proses menyerap, mengkaji, dan memanfaatkan informasi untuk merumuskan kesimpulan. Dengan kata lain, analisis dapat diartikan sebagai kegiatan menelaah suatu objek yang kemudian diikuti dengan pengolahan data sehingga menghasilkan suatu kesimpulan [10]. Selain itu, analisis juga merupakan upaya untuk menguraikan suatu permasalahan ke dalam bagian-bagian yang lebih kecil agar setiap bagian yang telah dipecah dapat terlihat dengan jelas, memiliki makna yang dapat dipahami, serta lebih mudah dimengerti [11].

## Iklim

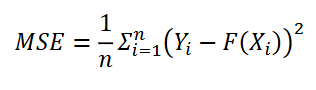
Iklim merupakan kondisi alam yang berubah secara mendadak dan tidak dapat diamati langsung oleh manusia. Iklim juga merujuk pada distribusi cuaca dalam periode waktu tertentu, baik harian, bulanan, maupun tahunan, dengan rata-rata serta kondisi ekstrem (maksimum dan minimum) dalam jangka waktu yang cukup lama [12]. Perubahan iklim adalah ancaman bagi peradaban manusia dan terkait dengan perubahan signifikan dalam pola cuaca global untuk waktu yang lama. Fenomena ini terutama disebabkan oleh peningkatan emisi gas rumah kaca seperti karbon .dioksida (CO2), .metana (CH4), dan nitrogen oksida (N2O). Ini disebabkan .oleh .aktivitas manusia seperti bahan bakar fosil, transformasi, dan praktik pertanian intensif [13].

## Gradient Boosting

Untuk regresi dan klasifikasi, gradien boosting adalah metode pembelajaran mesin dibangun menggunakan algoritma peningkatan gradien merupakan ansambel dari model prediktif yang lebih lemah; ini biasanya dibentuk dalam bentuk pohon keputusan. Pohon keputusan ini dilatih dengan memberikan nilai yang sama untuk setiap observasi. Setelah evaluasi awal, Nilai observasi yang sulit diklasifikasikan ditingkatkan, sedangkan nilai observasi yang mudah diklasifikasikan dan dikurangi [14].

## Mean Squared Error (MSE)

MSE merupakan .metrik yang menghitung rataan .dari selisih kuadrat antara nilai prediksi dan nilai aktual [15]. Rumus MSE adalah sebagai berikut:



Dimana :

Berikut adalah parafrase kalimat yang diminta tanpa mengurangi jumlah kata:

Dimana:

n : merupakan jumlah

keseluruhan observasi.

Yi : menunjukkan nilai aktual.

F(Xi) : merupakan nilai prediksi

yang dihasilkan oleh model

untuk observasi ke-i.

(Yi − F(Xi))² : merupakan perbedaan

absolut antara nilai aktual dengan nilai prediksi.

∑ : menyatakan jumlah

keseluruhan dari setiap perbedaan absolut.

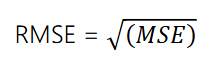
1/n : merupakan kebalikan dari jumlah keseluruhan

observasi, yang berfungsi sebagai rata-rata dari perbedaan absolut.

MSE menunjukkan seberapa efektif model dalam memprediksi nilai aktual, dan semakin kecil nilai MSE, semakin baik kinerja model dalam melakukan prediksi [15].

## Root Mean Squared Error (RMSE)

*Root Mean Squared Error* (RMSE) atau yang disebut salah satu .metrik yang dipakai untuk menilai .seberapa akurat model .dalam melakukan .prediksi. Metrik ini diperoleh dengan menghitung .akar kuadrat dari *Mean Squared Error* (MSE). Adapun rumus RMSE adalah sebagai berikut:



Dimana :

MSE : merupakan *Mean Squared Error*.

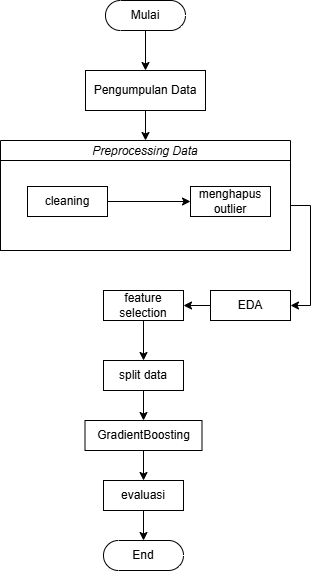
: merupakan fungsi akar kuadrat.

RMSE .kerap digunakan .dalam menilai kinerja model .prediksi karena memberikan gambaran .yangg intuitif serta ukuran yang mencerminkan besarnya kesalahann dalam hasiI prediksi. Nilai RMSE yang rendah atau mendekati .nol menandakan bahwa .hasil peramalan sagnat sesuai dengan data aktual. Meskipun RMSE umumnya bernilai positif, semakin kecil nilai RMSE, semakin tinggi tingkat akurasi prediksi yang dihasilkan [15].

## Machine Learning

*Machine Leaning* adalah bidang studi yang berfokus pada pengembangan kemampuan komputer untuk belajar dan berkembang dengan mentiru atau bahkan melampaui cara manusia dalam belajar [16]. Metode pembelajaran mesin otomatis dapat menyederhanakan proses ini dengan secara otomatis memilih model, mengoptimalkan hyperparameter, serta meminimaIkan keterlibatan manusia. Pendekatan ini juga berpotensi mengurangi beban kerja para ahIi radiologi dan patologi, sekaIigus menghemat waktu serta tenaga dalam menganalisis data pencitraan medis [17].

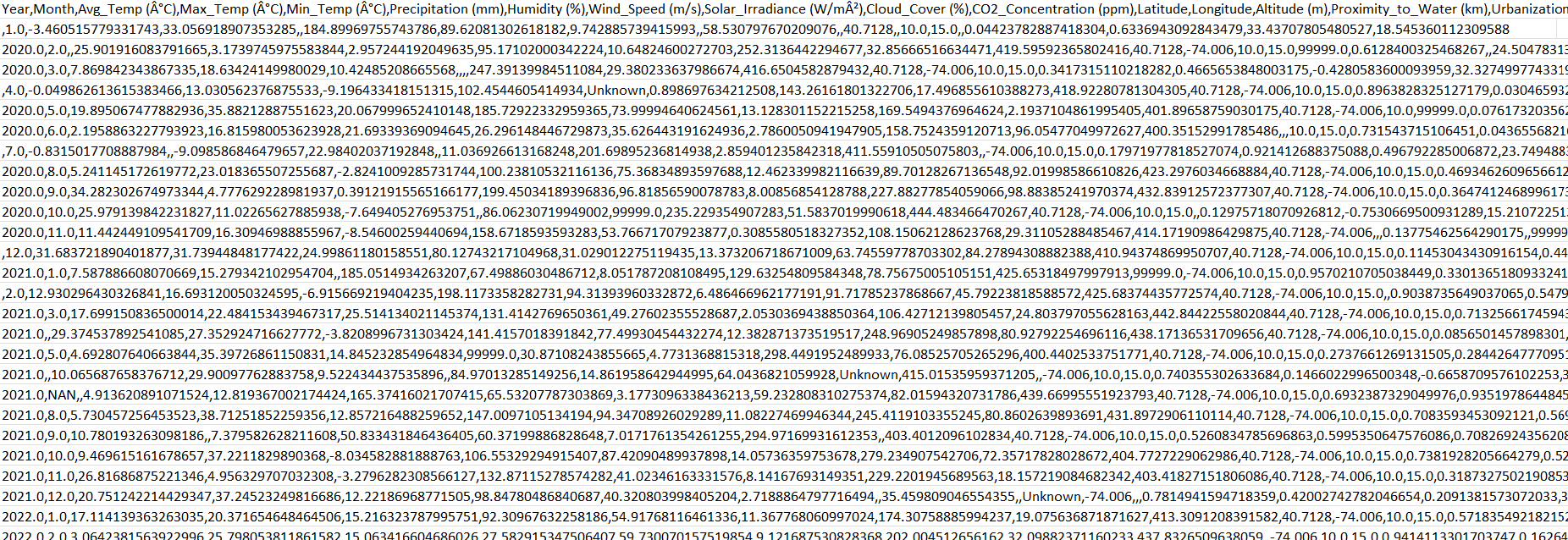
# METODE PENELITIAN



Gambar 1 Metodologi penelitian

## Pengumpulan Data

Selanjutnya, tahapan ini dilakukan dengan mengumpuIkan data yang berisi data iklim dari tahun 2020 sampai dengan tahun 2024, dan dalam penelitian ini dataset yang digunakan merupakan data yang didapat dari *platform kaggle* berjumlah 53 baris data dengan 20 kolom.



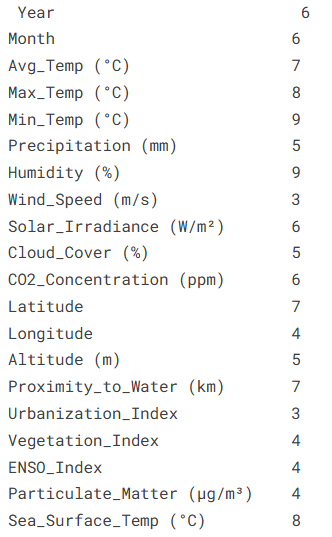
Gambar 2 Dataset iklim

## Preprocessing Data

Selanjutnya, *preprocessing* data .dilakukan untuk .mempersiapkan data sebelum masuk ke tahap implementasi algoritma dengan tujuan untuk membuat data menjadi bersih yang nantinya akan membuat hasil akurasi menjadi lebih baik. Berikut beberapa tahapan *preprocessing data.*

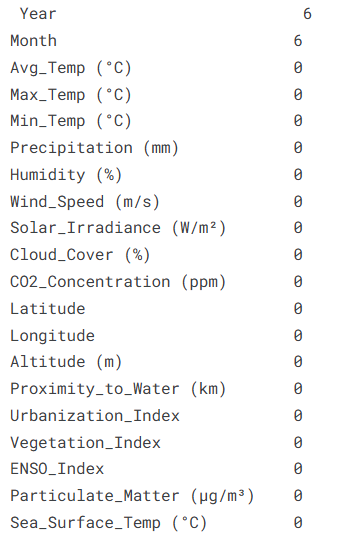
### Mengatasi Missing Values

Pada atribut yang dipilih tidak terdapat nilai yang kosong seperti pada Gambar 4 dan Gambar 5.



Gambar 4 *Missing values* pada dataset iklim

Pada Gambar 4 merupakan *missing values* pada dataset iklim, dimana pada tahapan ini data akan melewati proses pembersihan data dengan cara mengisi nilai *mean* pada dataset tanpa harus menghapusnya.

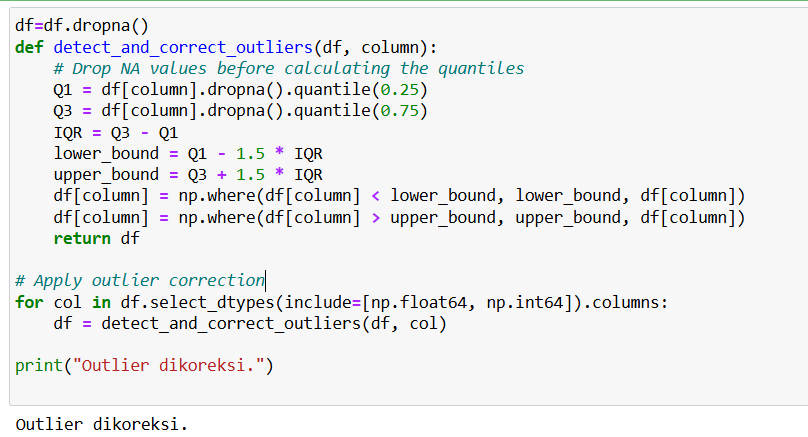


Gambar 5 Hasil *preprocessing data*

Pada Gambar 5 merupakan hasil *preprocessing data* dimana tidak terdapat nilai yang kosong.

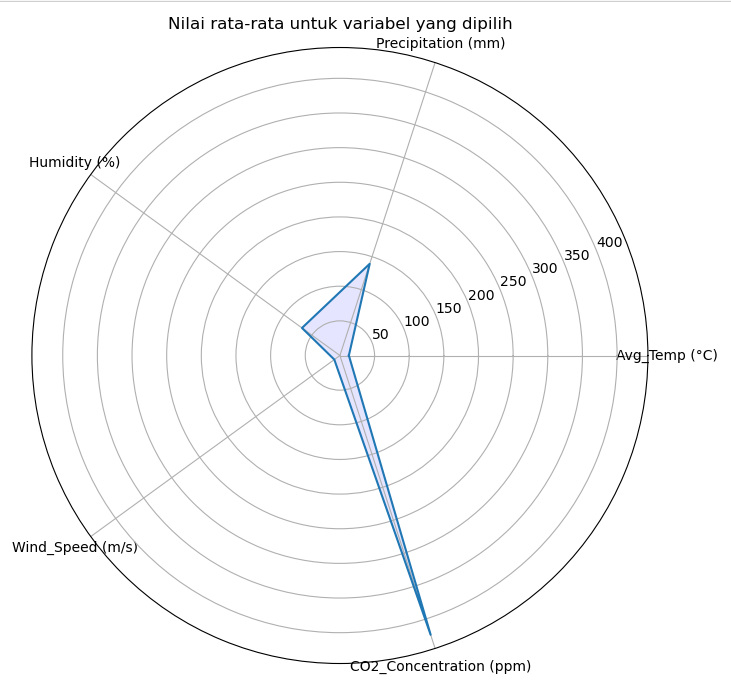
### Menghapus Nilai Outlier

Selanjutnya yaitu tahapan pemeriksaan outlier pada dataset. Adapun fungsi dari tahapan ini yaitu mengurangi dampak outlier ekstrem yang dapat menyebabkan model bias, menjaga distribusi data tetap seimbang dan meningkatkan performa model prediktif.

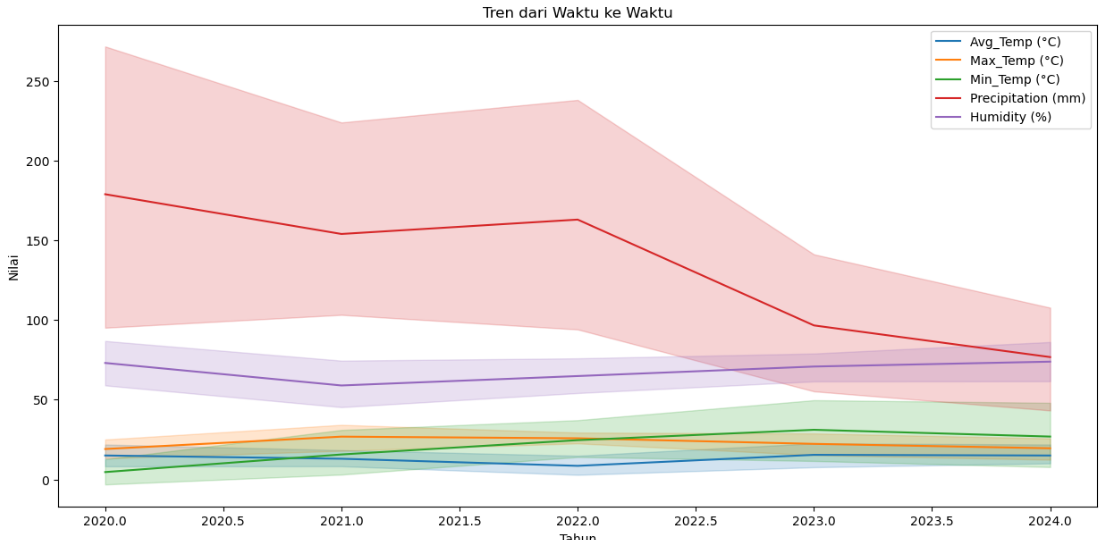


Gambar 6 Menghapus nilai *outlier*

### Exploratory Data Analysis (EDA)



Gambar 7 Nilai rata-rata variabel yang dipilih



Gambar 8 Tren dari waktu ke waktu

### Feature Selection

Adapun atribut yang digunakan pada penelitian ini yaitu variabel X nya “Year, Month” dan variabel Y nya “Avg\_Temp (°C), Precipitation (mm), Solar\_Irradiance (W/m²), Cloud\_Cover (%), CO2\_Concentration (ppm), Sea\_Surface\_Temp (°C)”. Korelasi antar variabel selanjutnya bisa dilihat pada Gambar 3.

### Splitting Data

*Splitting data* atau pemisaha data .dilakukan menjadi. dua bagian, yaitu data latih dan data uji [18]. Data latih berfungsi untuk melatih model, sementara data uji digunakan guna mengevaluasi performa model yang telah dilatih. Proses pemisahan data ini dilakukan dengan menerapkan proporsi 80:20.

## Implementasi Algoritma Gradient Boosting

Setelah tahapan *preprocessing data* dilakukan, algoritma *Gradient Boosting* diimplementasikan karena keunggulannya dalam menangani dataset kompleks dan tidak seimbang. Algoritma *gradient boosting* ini secara adaptif memprioritaskan data yang sulit diklasifikasikan dan mengelola ketidakseimbangan kelas melalui teknik penyeimbangan *loss* otomatis.

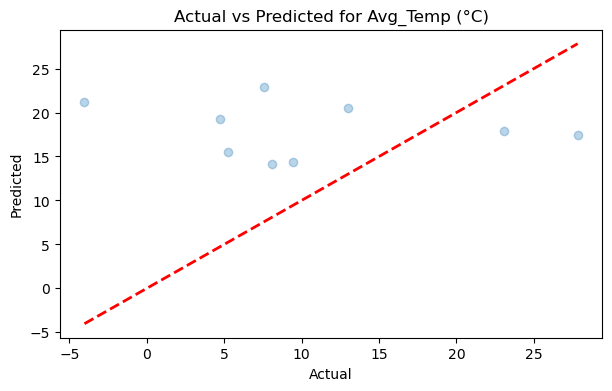
## Evaluasi

Terakhir, untuk mengukur kinerja model *Gradient Boosting* dalam memprediksi dan menganalisis pola perubahan iklim, digunakan metrik evaluasi RMSE dan MAE. RMSE mengukur seberapa besar kesalahan antara .nilai prediksi dan .nilai aktual dalam skala yang sama dengan unit data. MAE mengukur rata-rata perbedaan absolut antara prediksi dan nilai aktual.

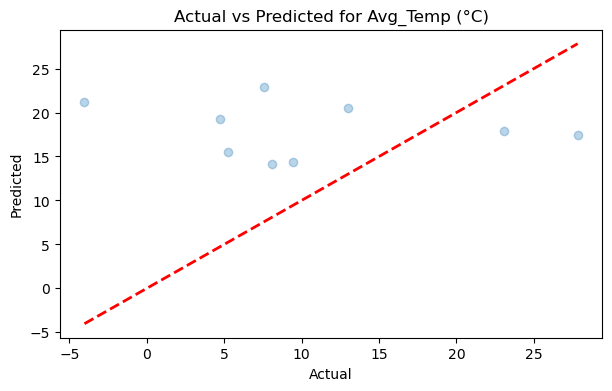
# HASIL DAN PEMBAHASAN

## Hasil Evaluasi Implementasi Algoritma Gradient Boosting

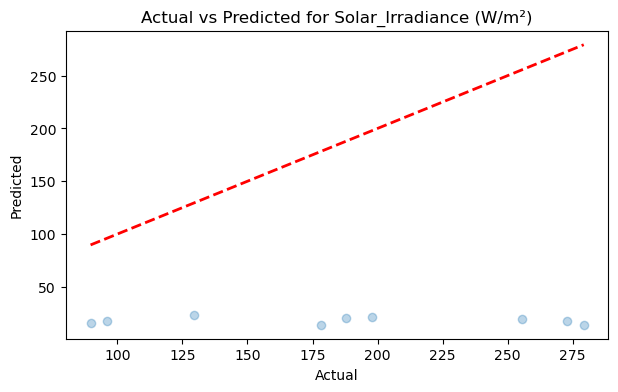
Algoritma yang digunakan pada penelitian ini .adalah *Gradient Boosting* untuk memprediksi dan menganalisis pola perubahan iklim. Hasil dari algoritma *Gradient Boosting* dalam pola perubahan iklim menggunakan metrik evaluasi RMSE dan MAE untuk setiap variabel dapat dilihat pada Gambar 9 sampai dengan Gambar 14.



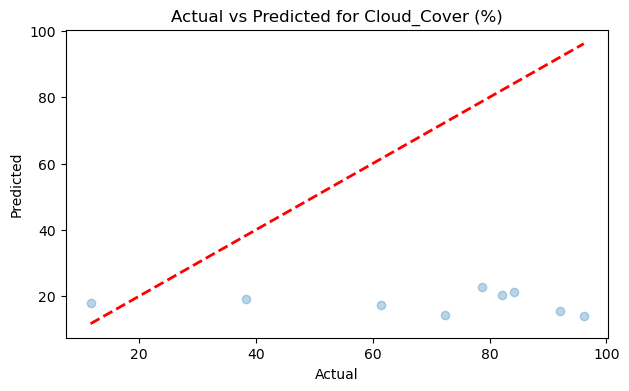
Gambar 9 Komparasi antara nilai aktual dan nilai prediksi Avg\_Temp



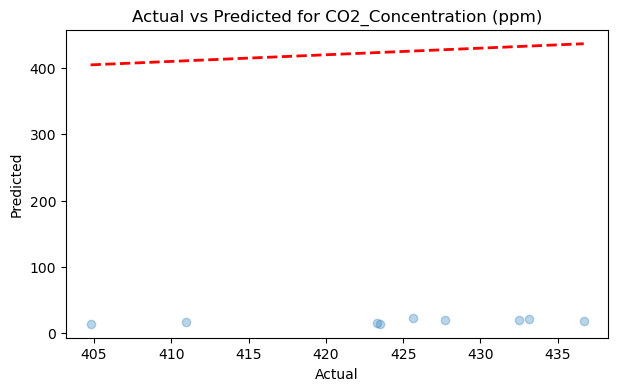
Gambar 10 Komparasi antara nilai aktual dan nilai prediksi Avg\_Temp (°C)



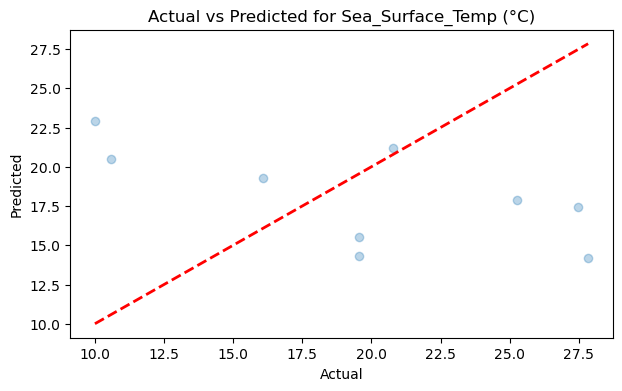
Gambar 11 Komparasi antara nilai aktual dan nilai prediksi Solar\_Irradiance (W/m²)



Gambar 12 Komparasi antara nilai aktual dan nilai prediksi Cloud\_Cover (%)



Gambar 13 Komparasi antara nilai aktual dan nilai prediksi CO2\_Concentration (ppm)



Gambar 14 Komparasi antara nilai aktual dan nilai prediksi Sea\_Surface\_Temp (°C)

Tabel 1 Hasil implementasi *gradient boosting* dengan RMSE dan MAE

|  |  |  |  |
| --- | --- | --- | --- |
| Target | Model | RMSE | MAE |
| Avg\_Temp (°C) | Gradient  Boosting  Regressor | 11.55 | 10.54 |
| Precipitation (mm) | 104.6 | 80.36 |
| Solar\_Irradiance (W/m²) | 68.29 | 62.00 |
| Cloud\_Cover (%) | 28.02 | 24.53 |
| CO2\_Concentration (ppm) | 10.43 | 8.92 |
| Sea\_Surface\_Temp (°C) | 8.57 | 7.42 |

Berdasarkan pada Tabel 1, pengujian yang dilakukan dengan RMSE dan MAE mendapatkan nilai yang bervariasi. Untuk **Avg Temp (°C),** RMSE nya adalah **11.55**. Untuk **Precipitation (mm),** RMSE nya adalah **104.6,** menunjukkan kesalahan prediksi curah hujan yang cukup besar. Untuk **CO2 Concentration (ppm),** MAE-nya adalah**8.92.** Untuk **Sea Surface Temp (°C),** MAE nya adalah **7.42,** menunjukkan kesalahan prediksi suhu permukaan laut yang relatif kecil. Model *Gradient Boosting* menunjukkan performa yang bervariasi tergantung pada variabel target. Performa terbaik terlihat pada prediksi CO2 Concentration (ppm) dan Sea Surface Temp (°C), sementara performa terburuk terlihat pada prediksi Precipitation (mm).

# KESIMPULAN

Bedasarkan pengujian yang dilakukan dengan menggunakan *Gradient Boosting* hasil prediksi dengan menggunakan metrik evaluasi RMSE dan MAE menunjukkan bahwa metode ini menghasilkan *error* yang yang kecil pada **CO2 Concentration (ppm),** denganRMSE nya 10.43, MAE nya adalah**8.92.** Untuk **Sea Surface Temp (°C),** RMSE nya 8.57, MAE nya adalah **7.42,** menunjukkan kesalahan prediksi suhu permukaan laut yang relatif kecil.

# UCAPAN TERIMA KASIH

Kami mengucapkan .terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan. dukungan dalam kelancaran penelitian ini, mulai dari tahap penulisan. hingga proses penerbitan. Ucapan terima kasih ini kami tujukan kepada orang tua kami, rekan-rekan, penerbit, serta banyak pihak lain yang tidak dapat kami sebutkan satu per satu. Selain itu, kami juga menyampaikan rasa terima kasih yang mendalam kepada tim Jurnal Informatika dan Teknik Elektro Terapan yang telah meluangkan waktu untuk membaca, mengevaluasi, serta memberikan kesempatan kepada kami untuk mempublikasikan hasil karya ilmiah ini. Semoga karya ini dapat menjadi sumber referensi yang berguna bagi para pembaca dan memberikan manfaat yang luas bagi masyarakat.

# DAFTAR PUSTAKA

[1] R. Ratnawati and R. N. Fauziah, “*Eyes on the Earth*: Situs Interaktif untuk Meningkatkan Kesadaran Perubahan Iklim Global,” *Bincang Sains dan Teknologi*, vol. 2, no. 02, pp. 82–89, Jun. 2023, doi: 10.56741/bst.v2i02.348.

[2] Agustina, N. A., Prasita, V. D., Kusuma, A., & Rosana, N. (2023). Pemanfaatan Sumberdaya Lahan Pesisir Berbasis Daya Dukung Lingkungan Dalam Menghadapi Perubahan Iklim Global.

[3] F.-K. Wang and T. Mamo, “*Gradient boosted regression model for the degradation analysis of prismatic cells*,” Computers & amp; Industrial Engineering, vol. 144, p. 106494, 2020.

[4] K. Kraugusteeliana, S. Muis, F. Nugroho, A. Karim, and Y. Siagian, “Data Mining Klasifikasi *Breast Cancer* Menerapkan Algoritma *Gradient Boosted Trees*,” J. MEDIA Inform. BUDI DARMA, vol. 7, no. 2, pp. 881–890, 2023.

[5] Giyanda, G., & Saidah, S. (2021). Auto Machine Learning dengan Menggunakan H2O AutoML untuk Prediksi Harga Bitcoin: Array. Jurnal Ilmiah KOMPUTASI, 20(2), 189-198.

[6] A. Amalia, M. Radhi, S. H. Sinurat, D. R. H. Sitompul, and E. Indra, “Prediksi Harga Mobil Menggunakan Algoritma Regressi Dengan Hyper-Parameter Tuning”, JUSIKOM PRIMA, vol. 4, no. 2, pp. 28 -32, Feb. 2022.

[7] Fitri, E, “Analisis Perbandingan Metode Regresi Linier, *Random Forest Regression* dan *Gradient Boosted Trees Regression Method* untuk Prediksi Harga Rumah”. Journal of Applied Computer Science and Technology, 4(1), 58-64, 2023.

[8] Setiyono A, Dharmawan A, Furqon Z, “Prediksi Faktor Penyebab Emisi CO2 pada Kendaraan Menggunakan Gradient Boosting Regression.” Jurnal Fokus Elektroda: Energi Listrik, Telekomunikasi, Komputer, Elektronika dan Kendali). 2024 Dec 1;9(4):243-7.

[9] Darmawan, Gerarldo Indra, Erwani Merry Sartika, Eric Chandra, Novie Theresia Br Pasaribu, and Heri Andrianto. "Penerapan Algoritma *Gradient Boosting* pada Sinyal EEG sebagai Pengendali Kursi Roda." ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika 12, no. 2 (2024): 541.

[10] T. Hidayat and D. S. Canta, “Analisis Kepuasan Pengguna Terhadap Penerapan Aplikasi Tokopedia dengan Menggunakan Metode TAM,” *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 2, p. 472, Apr. 2022, doi: 10.30865/jurikom.v9i2.4088.

[11] S. Surahmat, “Analisis Penerimaan Learning Manajemen Sistem Jurusan Manajemen Informatika Politeknik Negeri Sriwijaya Menggunakan *Technology Acceptance Model* (TAM),” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 1, Jan. 2024, doi: 10.23960/jitet.v12i1.3843.

[12] M. Amdah, Arfandi, and Nasrul, “Integrasi Pengetahuan Lokal Masyarakat Malino dalam Pengelolaan Sumber Daya Alam Hayati Sebagai Sumber Pembelajaran Non-Formal,” *Indonesian Journal of Fundamental and Applied Geography*, vol. 1, no. 2, pp. 15–18, 2024.

[13] M. Filonchyk, M. Peterson, L. Zhang, V. Hurynovich, and Y. He, “*Greenhouse gases emissions and global climate* *change*: *Examining the influence of* CO2, CH4, and N2O,” *Science of The Total Environment*, 2024.

[14] M. K. Dwipa Jaya, “Perbandingan Random Forest, Decision Tree, Gradient Boosting, Logistic Regression untuk Klasifikasi Penyakit Jantung,” JNATIA, vol. 2, pp. 1–5, 2023.

[15] I. Amansyah, J. Indra, E. Nurlaelasari, and A. Juwita, “Prediksi Penjualan Kendaraan Menggunakan Regresi Linear :  Studi Kasus pada Industri Otomotif di Indonesia,” *INNOVATIVE: Journal Of Social Science Research*, vol. 4, no. 4, pp. 1199–1216, 2024.

[16] M. M. Teye, “*Understanding Of Machine Learning With Deep Learning: Architectures, Workflow, Applications* *And Future Directions*,” Mdpi, Vol. 12, No. 5, Apr. 2023, Doi: Https://Doi.Org/10.3390/Computers12050091

[17] A. E. E. Rashed, A. M. Elmorsy, And A. E. M. Atwa, “*Comparative Evaluation Of Automated Machine Learning Techniques For Breast Cancer Diagnosis*,” *Biomed Signal Process Control*, Vol. 86, 2023, Doi: <Https://Doi.Org/10.1016/J.Bspc.2023.105016>.

[18] F. Atmojo, C. Nurlita, And Nurchima, “Analisis Pemanfaatan *Machine Learning* Guna Prediksi Indeks Pembangunan Manusia,” *Jurnal Sistem Informasi Dan Teknik Komputer*, Vol. 9, No. 2, 2024.