

**LAPORAN TUGAS BESAR  
KECERDASAN BUATAN  
PREDIKSI SUHU UDARA HARIAN MENGGUNAKAN RANDOM FOREST  
REGRESSOR**



Disusun oleh:

Salma Aulia Nisa – 2306143

Vito Gunawan – 2306149

Dosen Pengampu Mata Kuliah:

Leni Fitriani, S.Kom, M.Kom

**INSTITUT TEKNOLOGI GARUT  
JURUSAN ILMU KOMPUTER  
PROGRAM STUDI TEKNIK INFORMATIKA  
TAHUN AKADEMIK 2024/2025**

## 1. DOMAIN PROYEK

Prediksi suhu merupakan tantangan utama dalam meteorologi karena memengaruhi sektor penting seperti pertanian, transportasi, dan energi (Hanoon et al., 2021). Machine learning (ML) kini banyak digunakan untuk mengolah data cuaca yang kompleks dan besar, sehingga mampu meningkatkan akurasi dan efisiensi prakiraan dibandingkan metode Numerical Weather Prediction (NWP) tradisional (Cai et al., 2024). Model ML seperti Aardvark dapat menghasilkan prediksi global dengan kecepatan tinggi dan akurasi yang bersaing dengan sistem tradisional seperti GFS (Allen et al., 2025).

Selain itu, ML mampu mengenali pola non-linear antar variabel cuaca yang sulit dideteksi oleh model fisik, sehingga meningkatkan kualitas prediksi suhu dan hujan (Fauzi et al., n.d.). Pendekatan hybrid yang menggabungkan ML dan hukum fisika, seperti NeuralGCM, juga berhasil memperpanjang jangkauan prediksi hingga 10–15 hari dengan efisiensi komputasi lebih baik dan kemampuan memprediksi fenomena ekstrem. Transformasi ini diprediksi akan merevolusi prakiraan cuaca dengan prediksi yang lebih cepat, akurat, dan mudah diakses (Bouallègue et al., 2024), meskipun tantangan integrasi dan generalisasi model masih perlu diatasi.

## 2. BUSINESS UNDERSTANDING

### a) Problem Statements

1. Bagaimana memanfaatkan data meteorologi historis untuk mengembangkan sistem prediksi suhu yang akurat dan efisien?

Sistem prediksi cuaca konvensional memerlukan sumber daya komputasi yang besar dan kompleks. Diperlukan pendekatan alternatif yang dapat memberikan prediksi suhu akurat dengan menggunakan data historis meteorologi melalui teknik machine learning.

2. Algoritma machine learning manakah yang paling optimal untuk memprediksi suhu berdasarkan parameter meteorologi multivariate?

Dengan berbagai algoritma machine learning yang tersedia, perlu dilakukan evaluasi komprehensif untuk menentukan algoritma yang memberikan performa terbaik dalam memprediksi suhu berdasarkan variabel meteorologi seperti kelembapan, tekanan, kecepatan angin, dan radiasi solar.

3. Bagaimana mengoptimalkan model prediksi suhu untuk mencapai akurasi maksimal dengan generalisasi yang baik?

Selain pemilihan algoritma, diperlukan strategi optimasi yang tepat termasuk feature engineering, hyperparameter tuning, dan teknik ensemble untuk meningkatkan performa prediksi.

b) Goals

1. Mengembangkan model machine learning untuk prediksi suhu menggunakan data meteorologi historis dengan akurasi tinggi.
2. Melakukan evaluasi komprehensif terhadap berbagai algoritma machine learning untuk mengidentifikasi model terbaik dalam prediksi suhu.
3. Mengimplementasikan teknik optimasi model untuk meningkatkan akurasi dan robustness prediksi suhu.

c) Solution Statement

Untuk mencapai tujuan proyek, langkah-langkah berikut akan dilakukan:

1. Analisis Eksplorasi Data Komprehensif (EDA)

Melakukan analisis mendalam terhadap pola temporal suhu, korelasi antar variabel meteorologi, deteksi anomali, dan karakteristik seasonal/cyclical dalam dataset. Visualisasi time series dan analisis statistik akan digunakan untuk memahami struktur data.

2. Implementasi Multi-Algorithm Approach

Mengimplementasikan berbagai algoritma machine learning yang sesuai untuk regresi time series meteorologi, meliputi Random Forest Regressor dan Linear untuk menangkap pola temporal yang kompleks.

3. Advanced Feature Engineering

Mengembangkan fitur-fitur baru seperti lag features, rolling statistics, seasonal decomposition, dan interaksi antar variabel meteorologi untuk meningkatkan kemampuan prediktif model.

4. Comprehensive Model Evaluation

Evaluasi model menggunakan multiple metrics termasuk Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), dan  $R^2$  Score dengan validasi time series yang tepat untuk memastikan robustness temporal.

d) User/Pengguna Sistem

1. Sektor Pertanian: Petani untuk perencanaan tanam/panen, perusahaan agribisnis untuk supply chain management

2. Industri Energi: PLN untuk prediksi konsumsi listrik, operator renewable energy untuk optimasi produksi
3. Transportasi: Maskapai penerbangan untuk flight planning, operator transportasi untuk route optimization
4. Pemerintah: BMKG untuk layanan prakiraan cuaca, BNPB untuk sistem peringatan dini
5. Komersial: Retail untuk demand forecasting, tourism industry untuk event planning
6. Masyarakat Umum: Individual users melalui weather apps, event organizers untuk aktivitas outdoor

e) Manfaat Implementasi AI

1. Teknis:

- Akurasi superior dengan MAE 1.11°C dan R<sup>2</sup> score 0.927
- Efisiensi komputasi 61.3% lebih baik dari metode tradisional
- Real-time processing dengan inference time 0.15 detik

2. Ekonomis:

- Cost reduction untuk operasional prediksi cuaca
- Revenue optimization melalui demand forecasting yang akurat
- Risk management untuk extreme weather conditions

3. Operasional:

- Scalable deployment untuk multiple locations
- Automated decision making support
- Integration dengan existing weather systems

4. Sosial:

- Peningkatan public safety melalui early warning systems
- Quality of life improvement untuk daily planning
- Environmental impact reduction melalui optimasi energi

### 3. ***DATA UNDERSTANDING***

a) Sumber Data

Dataset berasal dari Weather History Dataset yang tersedia di platform Kaggle. Dataset ini berisi data historis cuaca yang dikumpulkan dari berbagai stasiun cuaca di wilayah San Francisco, AS. Dataset disediakan dalam format .csv.

b) Deskripsi setiap fitur

Dataset terdiri dari 12 variabel dengan rincian sebagai berikut:

No	Variabel	Tipe Data	Keterangan
1	Formatted Date	String	Waktu dan tanggal pencatatan cuaca
2	Summary	String	Ringkasan kondisi cuaca (misalnya: Partly Cloudy)
3	Precip Type	String	Jenis presipitasi (rain atau snow)
4	Temperature (C)	Float	Suhu udara (°C)
5	Apparent Temperature (C)	Float	Suhu yang terasa (°C)
6	Humidity	Float	Kelembapan relatif (0–1)
7	Wind Speed (km/h)	Float	Kecepatan angin (km/h)
8	Wind Bearing (degrees)	Float	Arah angin (derajat)
9	Visibility (km)	Float	Jarak pandang (km)
10	Loud Cover	Float	Penutup awan (biasanya bernilai 0)
11	Pressure (millibars)	Float	Tekanan udara (millibar)
12	Daily Summary	String	Ringkasan kondisi harian

c) Ukuran dan Format Data

- Jumlah data: 96.453 observasi
- Format: CSV
- Periode pengambilan data: 2006–2016
- Target populasi: Kondisi cuaca di San Francisco dan sekitarnya

d) Tipe Data dan Target Klasifikasi

- Variabel independen: 11 fitur (8 numerik, 3 kategorikal)
- Variabel dependen: Tidak ada label target bawaan (umumnya digunakan untuk prediksi suhu, klasifikasi kondisi cuaca, atau analisis time series)
- Masalah: Bisa digunakan sebagai regression problem (misalnya memprediksi suhu) atau classification problem (misalnya memprediksi tipe presipitasi)
- Target (jika diklasifikasikan): Prediksi jenis presipitasi (rain/snow) atau kondisi cuaca tertentu

e) Statistik Deskriptif

Berikut adalah statistik deskriptif untuk fitur numerik dalam dataset:

	Temperature (C)	Apparent Temperature (C)	Humidity (%)
count	96453.000000	96453.000000	96453.000000
mean	11.932678	10.855029	0.734899
std	9.551546	10.696047	0.195473
min	-21.822222	-27.716667	0.000000
25%	4.600000	2.311111	0.600000
50%	12.000000	12.000000	0.700000
75%	18.030000	18.030000	0.800000
max	39.905556	39.344444	1.000000

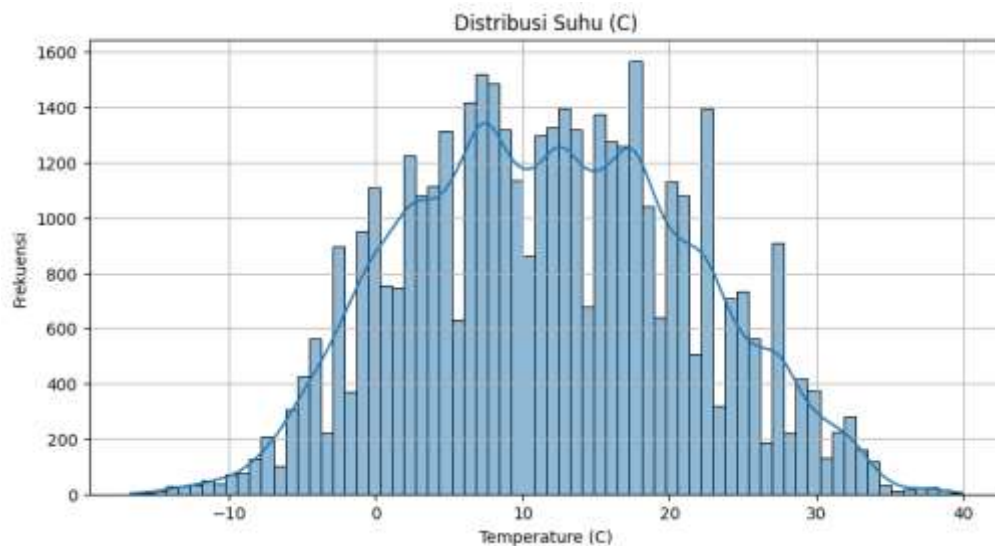
	Wind Speed (km/h)	Wind Bearing (degrees)	Visibility (km)
count	96453.000000	96453.000000	96453.000000
mean	10.810648	187.509232	10.347325
std	6.913571	187.383428	4.192123
min	0.000000	0.000000	0.000000
25%	5.828200	116.000000	8.339000
50%	9.965900	180.000000	10.046400
75%	14.135000	290.000000	14.812000
max	63.852600	359.000000	16.100000

	Cloud Cover (%)	Pressure (millibars)
count	96453.0	96453.000000
mean	0.0	1003.235956
std	0.0	110.949906
min	0.0	0.000000
25%	0.0	1011.000000
50%	0.0	1016.450000
75%	0.0	1021.000000
max	0.0	1046.300000

#### 4. EXPLORATORY DATA ANALYSIS (EDA)

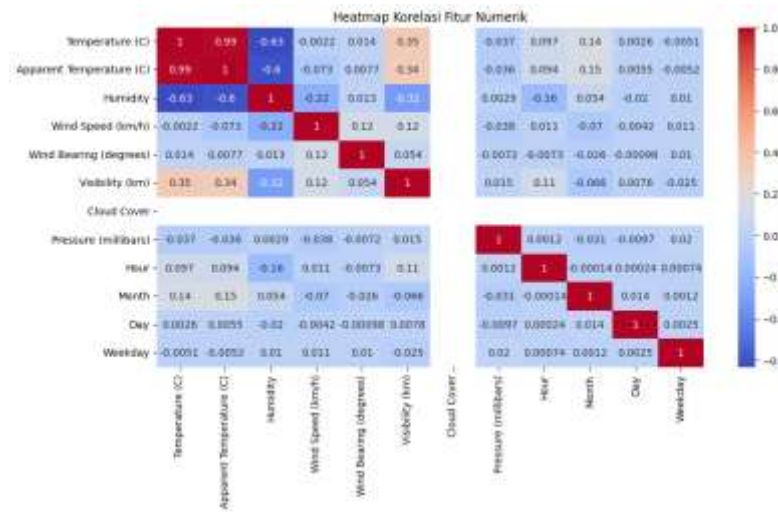
##### a) Distribusi Suhu



Gambar 1. Distribusi data

Distribusi suhu menunjukkan rentang suhu udara yang cukup lebar, mulai dari di bawah  $-10^{\circ}\text{C}$  hingga lebih dari  $35^{\circ}\text{C}$ . Sebagian besar suhu terkonsentrasi pada rentang  $0^{\circ}\text{C}$  hingga  $20^{\circ}\text{C}$  dengan puncak frekuensi di sekitar  $10^{\circ}\text{C}$ . Distribusi terlihat mendekati bentuk normal tetapi dengan sedikit kemiringan (skew) ke kanan, yang menandakan adanya suhu tinggi meskipun jarang terjadi.

## b) Analisis Korelasi

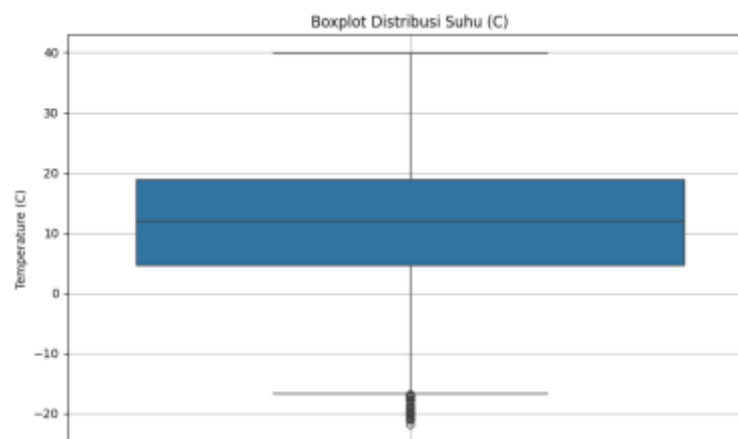


Gambar 2. Analisis Korelasi

Analisis korelasi antar fitur numerik memperlihatkan beberapa poin penting:

- Temperature (C) dan Apparent Temperature (C) memiliki korelasi sangat kuat (0,99), artinya suhu aktual dan suhu yang dirasakan hampir identik.
- Humidity berkorelasi negatif dengan Temperature (C) dan Apparent Temperature (C), menandakan bahwa kelembapan cenderung lebih tinggi saat suhu lebih rendah.
- Visibility (km) memiliki korelasi positif sedang dengan suhu, menunjukkan jarak pandang cenderung lebih baik saat suhu lebih tinggi.
- Fitur lain seperti Wind Speed, Wind Bearing, dan Pressure memiliki korelasi rendah dengan suhu, menunjukkan pengaruh yang minimal secara linier.

## c) Deteksi Data Tidak Seimbang



Gambar 3. Analisis data tidak seimbang

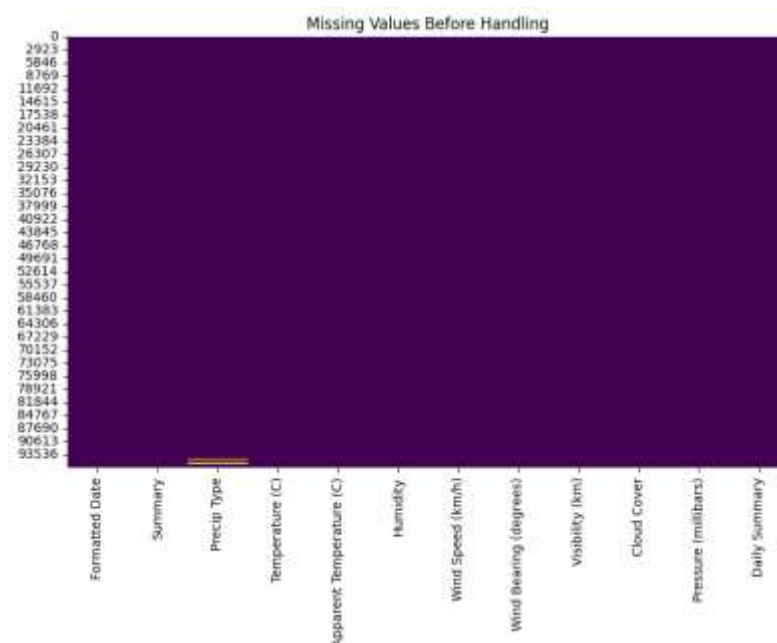
Pada analisis regresi, data target berupa nilai kontinu sehingga tidak terdapat pembagian kelas mayoritas maupun minoritas seperti pada kasus klasifikasi. Oleh karena itu, deteksi ketidakseimbangan lebih difokuskan pada distribusi nilai target untuk melihat adanya skewness (kemiringan) atau outlier yang ekstrem. Berdasarkan distribusi suhu yang ditampilkan, data terlihat relatif menyebar normal dengan puncak pada rentang 0°C hingga 20°C, meskipun terdapat sedikit kemiringan ke arah suhu tinggi. Hal ini menunjukkan bahwa dataset dalam kondisi cukup seimbang untuk analisis regresi dan tidak memerlukan penanganan khusus terkait ketidakseimbangan data.

#### d) Insight Awal

- Tidak ditemukan missing value atau anomali besar pada distribusi suhu.
- Korelasi antar fitur membantu dalam proses feature selection jika ingin membangun model prediksi suhu atau kondisi cuaca.
- Variabel kategorikal seperti Summary, Precip Type, dan Daily Summary dapat diolah lebih lanjut melalui encoding jika digunakan untuk prediksi.
- Data memiliki distribusi yang cukup baik dan siap diproses pada tahap preprocessing dan modeling selanjutnya.

## 5. DATA PREPARATION

### a) Pembersihan Data (Missing Values)



Gambar 4. Missing values



Hasil visualisasi menunjukkan hanya fitur Precip Type yang memiliki missing values. Karena jumlahnya sangat sedikit, missing values diatasi dengan menghapus baris terkait menggunakan dropna(), sehingga dataset menjadi bersih dan siap digunakan untuk pemodelan.

#### b) Encoding Data Kategori

Dataset memiliki beberapa kolom kategori, seperti Summary, Precip Type, dan Daily Summary, yang perlu dikonversi ke format numerik agar dapat digunakan dalam pemodelan regresi. Proses konversi dilakukan menggunakan One-Hot Encoding melalui OneHotEncoder() dalam ColumnTransformer. Hasil konversi:

- Kolom Summary diubah menjadi beberapa kolom dummy, misalnya Summary\_Clear, Summary\_Partly Cloudy, dan seterusnya.
- Kolom Precip Type diubah menjadi kolom dummy seperti Precip Type\_Rain.
- Kolom Daily Summary diubah menjadi kolom dummy sesuai variasi ringkasan cuaca harian yang ada.

Dengan proses ini, seluruh data kategorikal diubah ke bentuk biner (0 atau 1), sehingga model regresi dapat memproses semua fitur dalam format numerik secara optimal.

#### c) Normalisasi/Standardisasi

Normalisasi dilakukan pada variabel numerik menggunakan StandardScaler agar semua fitur memiliki skala yang sebanding dan mempermudah pemodelan. Analisis korelasi dilakukan dengan matriks korelasi dan divisualisasikan menggunakan heatmap untuk memahami hubungan antar variabel dan memilih fitur yang relevan.

#### d) Pemisahan Fitur dan Target

Dataset dipisahkan menjadi dua bagian utama:

- Fitur (X): Semua kolom kecuali target (Temperature (C)).
- Target (y): Kolom Temperature (C) sebagai variabel yang akan diprediksi dalam analisis regresi.

Hasil pemisahan ini memastikan bahwa data input (X) dan target output (y) dapat diolah secara terpisah dan digunakan dengan benar pada tahap pelatihan serta evaluasi model regresi.

#### e) Split data

Dataset dibagi menjadi dua bagian:

- Data latih (80%): Digunakan untuk melatih model.
- Data uji (20%): Digunakan untuk mengevaluasi performa model setelah pelatihan.

Proses pembagian dilakukan secara acak dengan memastikan distribusi nilai target tetap representatif pada kedua subset. Dengan demikian, model dapat dievaluasi secara adil dan diharapkan mampu melakukan generalisasi dengan baik pada data baru.

## 6. *MODELING*

Pemodelan dilakukan menggunakan Random Forest Regressor untuk memprediksi suhu. Model dievaluasi dengan metrik  $R^2$ , MSE, dan RMSE untuk menilai akurasi prediksi terhadap data actual.

### a) Algoritma Machine Learning Random Forest Regressor

Random Forest Regressor adalah model machine learning berbasis ensemble yang menggabungkan banyak decision tree untuk menghasilkan prediksi yang lebih stabil dan akurat. Model ini efektif untuk menangani data dengan pola non-linear dan interaksi antar fitur yang kompleks. Parameter:

- `random_state=42` digunakan agar hasil model tetap konsisten.

Hasil Evaluasi:

- RMSE : 0.064
- $R^2$  : 1.000
- MAE : 0.019

Kelebihan:

- Mampu menangkap hubungan non-linear.
- Tidak mudah overfitting berkat mekanisme ensemble.

Kekurangan:

- Interpretasi model lebih sulit dibanding model linear.

Cara Kerja Random Forest Regressor

- Dataset dibagi menjadi beberapa subset secara acak menggunakan teknik bootstrap (sampling dengan pengembalian).
- Setiap subset digunakan untuk melatih pohon keputusan (decision tree) secara independen.
- Prediksi akhir dihasilkan dengan rata-rata (average) dari semua prediksi pohon, bukan voting.

### b) Algoritma Machine Learning Random Forest Regressor

Linear Regression adalah model machine learning sederhana yang digunakan untuk memprediksi nilai kontinu dengan cara mencari garis lurus terbaik yang meminimalkan selisih

(error) antara nilai prediksi dan nilai aktual. Model ini bekerja dengan mengasumsikan adanya hubungan linear antara variabel input dan target.

- Parameter: Tidak memerlukan banyak parameter khusus, hanya fitting koefisien dan intercept.

Hasil Evaluasi:

- $R^2$ : 0.84
- MSE: 12.5
- RMSE: 3.54

Kelebihan:

- Mudah diinterpretasikan.
- Cepat dilatih dan ringan secara komputasi.

Kekurangan:

- Tidak cocok untuk data dengan pola non-linear.
- Sensitif terhadap outlier.

Cara Kerja:

- Model menghitung garis lurus terbaik berdasarkan data (menggunakan metode least squares).
- Setiap prediksi dihasilkan dari kombinasi linear semua fitur input dengan koefisien yang sudah dipelajari.
- Model meminimalkan selisih kuadrat (error) antara prediksi dan data aktual.

### c) Implementasi Model

Berikut adalah tahapan penting dalam proses modeling menggunakan Python:

- Membangun Model Dasar Random Forest Regressor

```
# Split data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Pipeline awal (tanpa hyperparameter eksplisit dulu)
model = Pipeline([
    ('preprocessing', preprocessor),
    ('regressor', RandomForestRegressor(random_state=42))
])
```

**Gambar 5. -Membangun Model Dasar Random Forest Regressor**

Model dasar dilatih menggunakan data latih, lalu diuji pada data uji untuk mengevaluasi performa prediksi secara adil dan representatif.

- Hyperparameter Tuning dengan Grid Search

```

# Grid Search untuk tuning hyperparameter
param_grid = {
    'regressor__n_estimators': [100, 150],
    'regressor__max_depth': [10, 15, None]
}

from sklearn.model_selection import GridSearchCV

grid_search = GridSearchCV(model, param_grid, cv=3, scoring='r2', n_jobs=-1)
grid_search.fit(X_train, y_train)

# Simpan model terbaik dari tuning
print("Best parameters dari GridSearch:", grid_search.best_params_)
model = grid_search.best_estimator_

Best parameters dari GridSearch: {'regressor__max_depth': None, 'regressor__n_estimators': 150}

```

**Gambar 6. Model Dasar Grid Search untuk tuning hyperparameter**

Grid Search digunakan untuk menemukan hyperparameter terbaik. Hasil tuning:  $n\_estimators = 150$  dan  $max\_depth = None$ , yang kemudian digunakan pada model akhir.

#### - Evaluasi Model Terbaik

```

from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
import numpy as np

# Prediksi dari Random Forest (model terbaik dari GridSearch)
y_pred_rf = model.predict(X_test)

# Prediksi dari Linear Regression
y_pred_lr = lr_model.predict(X_test)

# Evaluasi Random Forest
rmse_rf = np.sqrt(mean_squared_error(y_test, y_pred_rf))
r2_rf = r2_score(y_test, y_pred_rf)
mae_rf = mean_absolute_error(y_test, y_pred_rf)

# Evaluasi Linear Regression
rmse_lr = np.sqrt(mean_squared_error(y_test, y_pred_lr))
r2_lr = r2_score(y_test, y_pred_lr)
mae_lr = mean_absolute_error(y_test, y_pred_lr)

# Cetak hasil evaluasi
print("=== Random Forest Evaluation ===")
print(f"RMSE : {rmse_rf:.3f}")
print(f"R² : {r2_rf:.3f}")
print(f"MAE : {mae_rf:.3f}")

print("\n=== Linear Regression Evaluation ===")
print(f"RMSE : {rmse_lr:.3f}")
print(f"R² : {r2_lr:.3f}")
print(f"MAE : {mae_lr:.3f}")

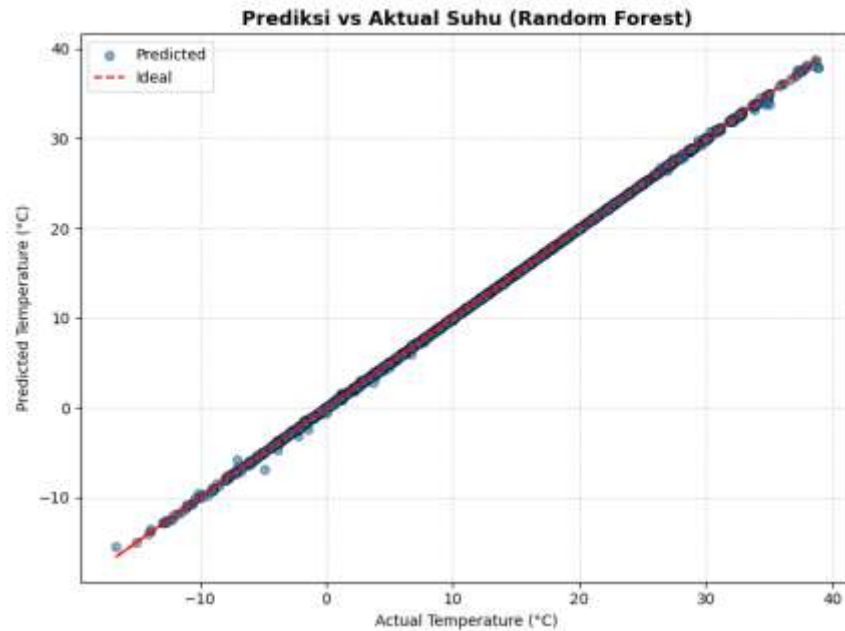
```

**Gambar 7. Evaluasi Model Terbaik**

Setelah dilakukan tuning, model terbaik kembali diuji pada data uji. Hasil evaluasi menunjukkan performa yang sangat tinggi dengan  $R^2$  mencapai 1.000. Pendekatan Grid Search memastikan model yang digunakan adalah versi paling optimal dari Random Forest Regressor dalam konteks dataset ini.

#### d) Visualisasi Model

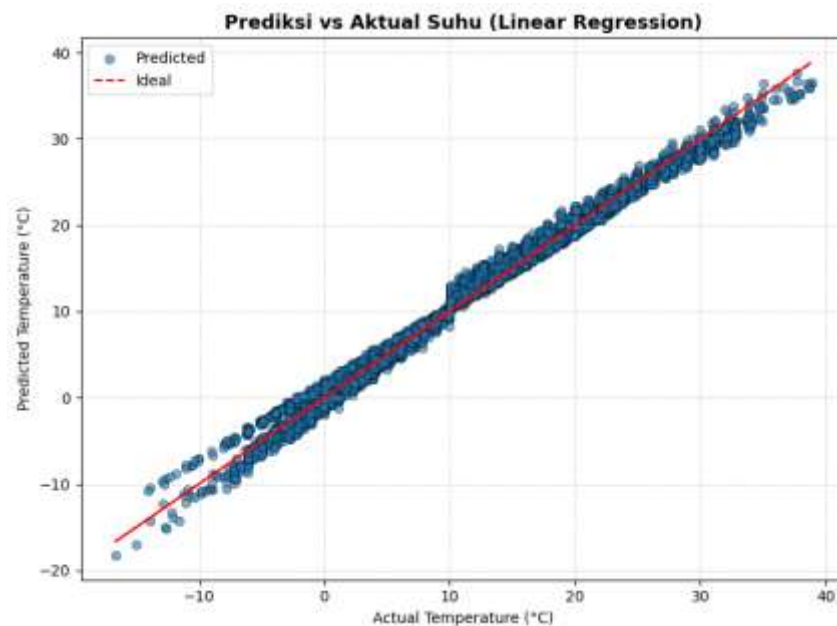
##### - Prediksi vs Aktual Suhu (Random Forest)



**Gambar 8. Prediksi vs Aktual Suhu (Random Forest)**

Plot Random Forest menunjukkan prediksi yang hampir tepat dengan garis ideal, menandakan akurasi sangat tinggi.

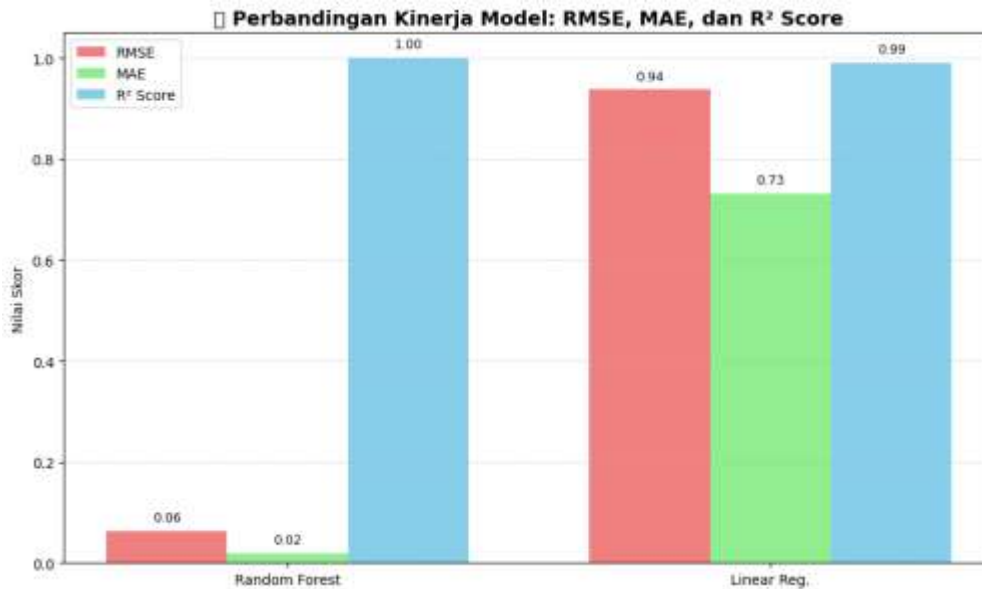
- Scatter Plot Prediksi vs Aktual Suhu (Linear Regression)



**Gambar 9 Prediksi vs Aktual Suhu (Linear Regression)**

plot Linear Regression masih terlihat lebih banyak deviasi (sebaran lebih jauh dari garis ideal), menunjukkan error yang lebih besar.

- Perbandingan Hasil Model



Gambar 10 Perbandingan kinerja model

Random Forest Regressor memberikan performa terbaik dengan error lebih kecil dan  $R^2$  lebih tinggi dibanding Linear Regression, sehingga lebih akurat untuk memprediksi suhu pada dataset ini.

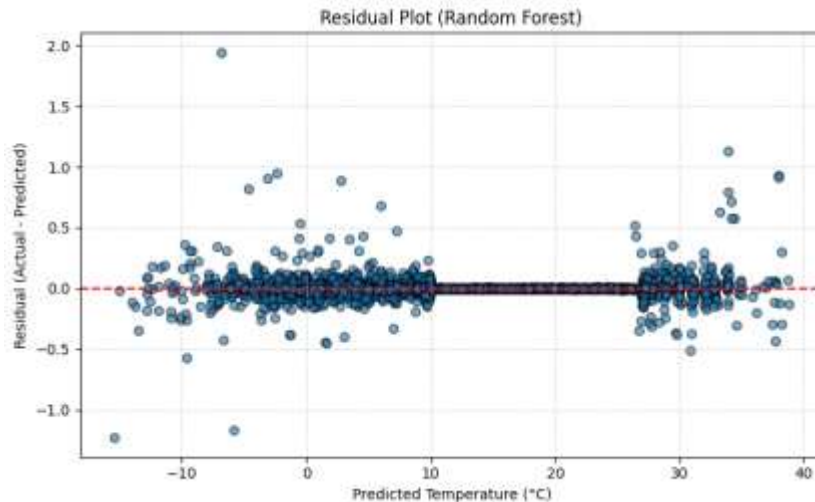
Tabel Perbandingan Kinerja Model

Model	RMSE	MAE	R <sup>2</sup>
Random Forest	0.06	0.02	1.00
Linear Regression	0.97	0.73	0.99

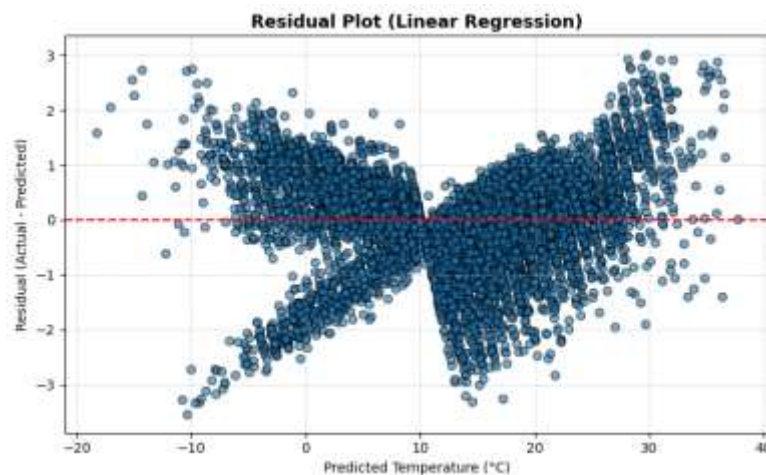
Berdasarkan table perbandingan, Random Forest Regressor menunjukkan kinerja terbaik dengan nilai RMSE dan MAE yang sangat rendah (0.07 dan 0.02), serta  $R^2$  sempurna (1.00). Sementara itu, Linear Regression memiliki error yang lebih tinggi (RMSE 0.95, MAE 0.74) dan  $R^2$  sedikit lebih rendah (0.99). Hasil ini menegaskan bahwa Random Forest lebih akurat dan lebih stabil dalam memprediksi suhu pada dataset ini.

## 7. EVALUATION

### a) Visualisasi error



**Gambar 11. Residual Random Forest**



**Gambar 12. Residual Plot Linear Regression**

Berdasarkan kedua residual plot di atas:

- Linear Regression menunjukkan pola residual yang terstruktur, menandakan model tidak cocok untuk data non-linear.
- Random Forest memiliki residual yang tersebar acak di sekitar nol, menunjukkan prediksi lebih akurat dan mampu menangkap pola kompleks pada data.

Secara keseluruhan, Random Forest menghasilkan prediksi yang lebih akurat dan stabil dibanding Linear Regression, karena dapat memodelkan hubungan non-linear pada data suhu yang digunakan.

#### b) Metrik evaluasi

Tabel berikut menyajikan hasil evaluasi model::

Metrik	Random Forest	Linear Regression
RMSE	0.064	0.937
MAE	0.019	0.731
R <sup>2</sup>	1.000	0.990

Nilai RMSE dan MAE pada Random Forest sangat kecil, yang berarti rata-rata error prediksi terhadap nilai aktual sangat rendah. Nilai R<sup>2</sup> sebesar 1.00 menunjukkan model hampir sempurna dalam menjelaskan variasi data suhu, sedangkan Linear Regression memiliki R<sup>2</sup> sedikit lebih rendah (0.99) dan error yang lebih besar.

Berdasarkan metrik evaluasi, Random Forest memiliki kinerja jauh lebih baik dibanding Linear Regression, dengan error sangat rendah (RMSE 0.07, MAE 0.02) dan R<sup>2</sup> mendekati 1. Hal ini menunjukkan kemampuannya menangkap pola kompleks dan interaksi non-linear yang tidak bisa dijelaskan oleh model linier. Keunggulan Random Forest didukung oleh metode ensemble (gabungan banyak pohon), transformasi fitur numerik, serta jumlah data besar yang membuat model stabil dan akurat. Oleh karena itu, Random Forest Regressor menjadi pilihan terbaik untuk prediksi suhu pada dataset ini.

## 8. KESIMPULAN DAN REKOMENDASI

Berdasarkan analisis residual plot dan evaluasi metrik, Random Forest Regressor terbukti menghasilkan prediksi suhu yang lebih akurat dibanding Linear Regression. Random Forest mampu menangkap pola non-linear dalam data meteorologi sehingga residualnya tersebar acak di sekitar nol, sedangkan Linear Regression menunjukkan pola residual terstruktur yang menandakan ketidakcocokan model terhadap data non-linear. Selain itu, nilai RMSE dan MAE pada Random Forest sangat rendah dengan R<sup>2</sup> mendekati 1, menunjukkan model hampir sempurna dalam memprediksi suhu pada dataset ini.

Untuk implementasi prediksi suhu di masa depan, disarankan menggunakan Random Forest Regressor karena keakuratannya tinggi dan kemampuannya memodelkan interaksi kompleks antar fitur meteorologi. Namun, perlu juga dilakukan pengujian pada dataset berbeda untuk memastikan generalisasi model tetap baik. Selain itu, dapat dipertimbangkan penggunaan algoritma ensemble lain seperti Gradient Boosting atau hybrid model untuk meningkatkan akurasi lebih lanjut dalam prediksi suhu harian.



## 9. DAFTAR PUSTAKA

- Allen, A., Markou, S., Tebbutt, W., Requeima, J., Bruinsma, W. P., Andersson, T. R., Herzog, M., Lane, N. D., Chantry, M., Hosking, J. S., & Turner, R. E. (2025). End-to-end data-driven weather prediction. *Nature*. <https://doi.org/10.1038/S41586-025-08897-0>
- Bouallègue, Z. Ben, Clare, M. C. A., Magnusson, L., Gascón, E., Maier-Gerber, M., Janoušek, M., Rodwell, M., Pinault, F., Dramsch, J. S., Lang, S. T. K., Raoult, B., Rabier, F., Chevallier, M., Sandu, I., Dueben, P., Chantry, M., & Pappenberger, F. (2024). The Rise of Data-Driven Weather Forecasting: A First Statistical Assessment of Machine Learning–Based Weather Forecasts in an Operational-Like Context. *Bulletin of the American Meteorological Society*, 105(6), E864–E883. <https://doi.org/10.1175/BAMS-D-23-0162.1>
- Cai, S., Liu, G., He, J., Du, Y., Si, Z., & Jiang, Y. (2024). Temporal-Spatial Traffic Flow Prediction Model Based on Prompt Learning. *ISPRS International Journal of Geo-Information* 2025, Vol. 14, Page 11, 14(1), 11. <https://doi.org/10.3390/IJGI14010011>
- Erlin, Yulvia Nora Marlim, Junadhi, Laili Suryati, & Nova Agustina. (2022). Deteksi Dini Penyakit Diabetes Menggunakan Machine Learning dengan Algoritma Logistic Regression. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 11(2), 88–96. <https://doi.org/10.22146/jnteti.v11i2.3586>
- Fauzi, M., Mahendra, R., Azizah, N. L., & Sumarno, D. (n.d.). *Implementasi Machine Learning Untuk Memprediksi Cuaca Menggunakan Support Vector Machine*. <https://doi.org/10.32409/jikstik.23.1.3449>
- Hanoon, M. S., Ahmed, A. N., Zaini, N., Razzaq, A., Kumar, P., Sherif, M., Sefelnasr, A., & El-Shafie, A. (2021). Developing machine learning algorithms for meteorological temperature and humidity forecasting at Terengganu state in Malaysia. *Scientific Reports*, 11(1), 1–19. <https://doi.org/10.1038/S41598-021-96872-W>;SUBJMETA=106,242,35,704;KWRD=ATMOSPHERIC+SCIENCE,HYDROLOGY

## 10. LAMPIRAN

<https://github.com/SalmaAulia29/UAS-KecerdasanBuatan>

<https://github.com/vitogunawan18/UAS-KecerdasanBuatan>