IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 5, No. 1 Januari 2026
https://subset.id/index.php/IJCSR

Analisis Komparatif Algoritma Random Forest dan Multiple Linear Regression untuk Prediksi Suhu Udara dengan Rekayasa Fitur Siklik

Haryoko[1, a)], Lilis Dwi Farida*[2,b)], Afrig Aminuddin[3, c)], Bahrun Ghozali[4, d)]

[1)]*Teknologi Informasi,*
[2)] *Manajemen Informatika,*
[3)] *Sistem Informasi,*
[4)] *Teknik Informatika*
Fakultas Ilmu Komputer
Universitas Amikom Yogyakarta, Jl. Padjajaran, Ring Road Utara, Daerah Istimewa Yogyakarta 55283, Indonesia

Author Emails
[a)]haryoko.amikom.ac.id
[b)] Corresponding author: rida@amikom.ac.id
[c)]afrig@amikom.ac.id
[d)]oza@amikom.ac.id

Abstract. *Temperature prediction represents a significant challenge in meteorology and precision agriculture due to the highly fluctuating and non-linear characteristics of atmospheric data. This study aims to conduct a comparative analysis of Machine Learning algorithms for daily air temperature estimation by evaluating the performance of Multiple Linear Regression as a baseline model and Random Forest Regressor as the primary model. A comprehensive feature engineering process was applied, particularly transforming wind direction variables into continuous vector components and extracting temporal features into cyclical components to capture seasonal patterns. Based on evaluations using historical weather datasets, the Random Forest model demonstrated superior performance over Multiple Linear Regression, achieving a coefficient of determination (R^2) of 0.990, a Mean Absolute Error (MAE) of 0.74°C, and a Root Mean Squared Error (RMSE) of 0.95. These findings indicate that the combination of ensemble learning algorithms with appropriate cyclical feature handling is more effective in capturing complex weather variable interactions compared to conventional linear methods.*

Keywords**:** *Weather Prediction, Random Forest, Linear Regression, Feature Engineering, Cyclic Features.*

Abstraksi. Prediksi temperatur cuaca merupakan tantangan signifikan dalam bidang meteorologi dan pertanian presisi dikarenakan karakteristik data atmosfer yang sangat fluktuatif dan non-linear. Penelitian ini bertujuan untuk melakukan analisis komparatif kinerja algoritma Machine Learning dalam mengestimasi suhu udara harian, dengan membandingkan model Multiple Linear Regression sebagai baseline dan Random Forest Regressor sebagai model utama. Eksperimen melibatkan penerapan teknik rekayasa fitur (feature engineering) yang komprehensif, khususnya transformasi variabel arah angin menjadi komponen vektor kontinu serta ekstraksi fitur waktu menjadi komponen siklik untuk menangkap pola musiman. Berdasarkan hasil evaluasi menggunakan dataset historis cuaca, model Random Forest menunjukkan performa yang lebih superior dibandingkan Multiple Linear Regression, dengan capaian nilai Koefisien Determinasi (R^2) sebesar 0.990, Mean Absolute Error (MAE) sebesar 0.74°C, dan Root Mean Squared Error (RMSE) sebesar 0.95. Hasil analisis membuktikan bahwa pendekatan ensemble learning dengan penanganan fitur siklik jauh lebih efektif dalam memetakan kompleksitas interaksi variabel cuaca dibandingkan metode linear konvensional.

Kata Kunci: Prediksi Cuaca, Random Forest, Regresi Linear, Rekayasa Fitur, Fitur Siklik.

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 5, No. 1 Januari 2026
https://subset.id/index.php/IJCSR

PENDAHULUAN

Cuaca merupakan fenomena atmosfer yang dinamis dan kompleks yang memengaruhi berbagai aspek kehidupan manusia, mulai dari sektor pertanian, transportasi, hingga perencanaan energi [1]. Perubahan parameter meteorologi seperti suhu, kelembaban, dan tekanan udara terjadi secara fluktuatif dan sering kali dipengaruhi oleh interaksi non-linear yang sulit diprediksi menggunakan metode konvensional. Ketidakpastian ini menuntut adanya sistem peramalan yang tidak hanya akurat tetapi juga mampu beradaptasi dengan pola data yang rumit. Dalam beberapa tahun terakhir, pendekatan berbasis data (*data-driven*) menggunakan algoritma *Machine Learning* telah menjadi solusi alternatif [2][3] yang menjanjikan untuk mengatasi keterbatasan model numerik fisik yang sering kali membutuhkan komputasi berat.

Permasalahan utama dalam prediksi suhu cuaca adalah bagaimana menangkap pola siklus harian dan musiman serta hubungan antar variabel yang tidak selalu linear [4]. Metode statistik klasik seperti Regresi Linear sering kali gagal memodelkan kompleksitas ini secara optimal karena asumsi linearitas yang kaku. Selain itu, data mentah meteorologi, khususnya arah angin yang bersifat siklik dan data waktu yang berurutan, memerlukan perlakuan khusus agar dapat diproses dengan benar oleh algoritma. Tanpa teknik pra-pemrosesan (*preprocessing*) dan rekayasa fitur (*feature engineering*) yang tepat, model cerdas sekalipun dapat menghasilkan bias prediksi yang signifikan, terutama pada kondisi cuaca ekstrem atau anomali.

Penelitian ini bertujuan untuk mengembangkan dan membandingkan kinerja model prediksi temperatur cuaca menggunakan algoritma *Multiple Linear Regression* [1] dan *Random Forest Regressor* [5][6]. Fokus utama penelitian terletak pada penerapan teknik rekayasa fitur, yaitu transformasi variabel arah angin menjadi komponen vektor kontinu serta ekstraksi fitur waktu menjadi komponen siklik, guna meningkatkan kemampuan model dalam memahami pola data. Algoritma *Random Forest* dipilih sebagai metode utama karena kemampuannya dalam menangani non-linearitas dan ketahanannya terhadap overfitting melalui mekanisme *ensemble learning*. Melalui pendekatan ini, diharapkan dapat dihasilkan model estimasi suhu yang memiliki akurasi tinggi dan tingkat kesalahan yang rendah.

Kontribusi penelitian ini diharapkan dapat memberikan wawasan baru mengenai pentingnya tahapan rekayasa fitur dalam meningkatkan performa model *Machine Learning* pada data meteorologi. Penggunaan *Random Forest* yang dibandingkan dengan metode linear diharapkan dapat menjadi rujukan bagi pengembangan sistem implementasi ramalan cuaca di masa mendatang. Selain itu, penelitian ini akan mengusulkan model yang memiliki potensi implementasi praktis untuk mendukung pengambilan keputusan di bidang-bidang yang sensitif terhadap perubahan suhu, menyediakan alat bantu prediksi yang efisien tanpa memerlukan infrastruktur superkomputer yang mahal.

TINJAUAN PUSTAKA

Perkembangan teknologi prediksi cuaca saat ini sedang mengalami pergeseran paradigma yang signifikan dari model numerik murni (*Numerical Weather Prediction* atau NWP) menuju integrasi kecerdasan buatan (*Artificial Intelligence*). Studi literatur menunjukkan bahwa meskipun model NWP berbasis fisika telah menjadi standar emas, model-model ini masih memiliki keterbatasan komputasi dan kerentanan terhadap bias parameter fisik tertentu. Sebagai contoh, survei teknis terbaru menyoroti potensi besar AI dan *Deep Learning* dalam melengkapi atau bahkan menggantikan komponen tertentu dari NWP untuk meningkatkan efisiensi dan akurasi prediksi cuaca dan iklim [7]. Salah satu kelemahan kritis NWP konvensional adalah pengabaian interaksi aerosol-radiasi yang kompleks, yang terbukti menjadi faktor utama penyebab bias pada prakiraan suhu udara global; integrasi data aerosol ke dalam model terbukti dapat mereduksi kesalahan prediksi tersebut secara signifikan [8].

Dalam konteks pertanian presisi, kebutuhan akan data cuaca dengan resolusi spasial yang sangat tinggi menjadi semakin mendesak, terutama untuk lahan dengan topografi kompleks yang tidak dapat ditangkap oleh model skala grid standar. Penelitian terkini telah berhasil mengembangkan metode downscaling untuk menghasilkan data suhu udara permukaan (*Surface Air Temperature*) dengan resolusi 50 meter, yang mengoreksi output model cuaca numerik menggunakan teknik machine learning dan skala pendinginan radiatif untuk manajemen pertanian yang lebih presisi [9]. Sejalan dengan hal tersebut, pendekatan berbasis *Time-Series Mixer* juga mulai diterapkan untuk mendeteksi kejadian cuaca berbahaya di tingkat lokal, memungkinkan mitigasi risiko yang lebih baik bagi sektor pertanian melalui prediksi parameter lingkungan yang spesifik dan terlokalisasi [10].

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 5, No. 1 Januari 2026
https://subset.id/index.php/IJCSR

Selain prediksi parameter dasar, pemanfaatan data cuaca historis dan prediksi juga telah berkembang menjadi alat bantu pengambilan keputusan agronomis yang canggih, seperti pemodelan fenologi tanaman. Pengembangan paket perangkat lunak berbasis R seperti DyMEP (*Dynamic Multi-Environmental Phenology*) memungkinkan peneliti untuk memprediksi tahapan fenologi pada berbagai jenis tanaman (seperti gandum dan jelai) dengan memanfaatkan variabel lingkungan sebagai kovariat utama [11]. Pendekatan semacam ini menjembatani kesenjangan antara model pertumbuhan tanaman yang terlalu kompleks dengan model derajat-hari (*degree-day*) yang terlalu sederhana, memberikan fleksibilitas yang dibutuhkan untuk adaptasi terhadap variabilitas iklim regional.

Di sektor energi dan lingkungan binaan (*built environment*), akurasi prediksi cuaca memainkan peran vital dalam estimasi beban energi dan potensi pembangkitan energi terbarukan. Untuk memprediksi konsumsi energi bangunan sejak tahap desain, teknik clustering cuaca berbasis k-means telah digunakan untuk memilih hari-hari representatif, yang kemudian diproses menggunakan model pengganti (*surrogate model*) berbasis *machine learning* guna mempercepat simulasi tanpa mengorbankan akurasi [12]. Sementara itu, untuk pembangkitan listrik tenaga surya (*fotovoltaik*), tantangan ketidakpastian iradiasi matahari diatasi dengan menggabungkan prediksi ansambel NWP dengan rantai model fisik ansambel, yang menghasilkan prakiraan daya probabilistik yang lebih andal dibandingkan metode deterministik tunggal [13].

Penelitian juga merambah pada estimasi parameter meteorologi yang lebih spesifik dan kompleks untuk aplikasi khusus, seperti hidrologi dan transportasi. Dalam pemodelan evapotranspirasi, misalnya, pendekatan *Maximum Entropy Production* (MEP) telah diterapkan untuk memprediksi kelembaban spesifik permukaan tanah (*land-surface specific humidity*) dengan memanfaatkan suhu radiatif dan data cuaca ambien, mengatasi tantangan kelangkaan data pengamatan langsung di lapangan [14]. Di sisi lain, untuk aplikasi kendaraan otonom dan keselamatan berkendara, model prediksi intensitas presipitasi yang "dirasakan" (*perceived precipitation intensity*) telah dikembangkan dengan memperhitungkan kecepatan kendaraan (pengamatan dinamis) dibandingkan pengamatan statis, yang divalidasi melalui uji terowongan angin dan lapangan [15].

Terakhir, masa depan sistem prediksi cuaca operasional bergerak menuju pemanfaatan *High-Performance Computing* (HPC) dan konsep Digital Twins. Inisiatif skala besar seperti "*Destination Earth*" yang dikembangkan oleh ECMWF bertujuan menciptakan replika digital bumi yang sangat presisi untuk memantau dan memprediksi fenomena alam serta aktivitas manusia dengan detail yang belum pernah ada sebelumnya [16]. Transformasi infrastruktur ini menandai era baru di mana integrasi data besar, komputasi awan, dan algoritma cerdas bekerja secara sinergis untuk menghasilkan sistem peringatan dini dan pemantauan iklim yang jauh lebih responsif dan akurat.

METODE PENELITIAN

Penelitian ini dilaksanakan melalui tahapan sistematis yang dilakukan untuk membangun model prediksi temperatur cuaca. Pendekatan yang digunakan dalam penelitian ini adalah eksperimen kuantitatif berbasis data (*data-driven*) dengan menerapkan algoritma *Machine Learning*. Secara garis besar, metodologi ini mencakup proses pengumpulan data dari dataset historis, teknik pra-pemrosesan untuk meningkatkan kualitas data, implementasi algoritma Linear Regression dan *Random Forest*, hingga evaluasi kinerja model menggunakan metrik statistik standar. Uraian lengkap mengenai objek penelitian, alat yang digunakan, serta alur kerja sistem yang dijelaskan sebagai berikut.

Objek dan Sumber Data

Objek utama dalam penelitian ini adalah data meteorologi historis yang merekam kondisi cuaca per jam dalam rentang waktu tertentu. Sumber data diperoleh dari penyedia dataset publik Kaggle dengan nama berkas weatherHistory.csv. Dataset ini dipilih sebagai objek penelitian karena memiliki cakupan parameter atmosfer yang komprehensif serta volume data yang memadai untuk kebutuhan pelatihan model Machine Learning. Data tersebut tersedia dalam format terstruktur *Comma Separated Values* (CSV), yang memungkinkan pengolahan data secara efisien menggunakan pustaka manipulasi data seperti Pandas.

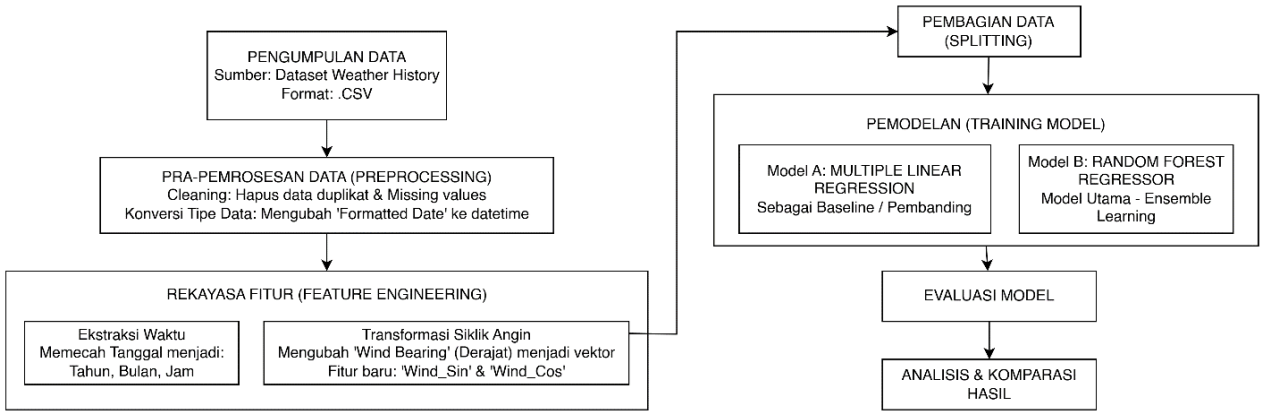
Secara spesifik, dataset ini terdiri dari satu variabel dependen (target variable) dan sejumlah variabel independen (*features*). Variabel target yang akan diprediksi adalah Temperature (Suhu dalam satuan derajat Celcius). Adapun variabel independen yang digunakan sebagai prediktor meliputi *Humidity* (Kelembaban), *Wind Speed* (Kecepatan

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 5, No. 1 Januari 2026
https://subset.id/index.php/IJCSR

Angin), *Wind Bearing* (Arah Angin), *Visibility* (Jarak Pandang), *Pressure* (Tekanan Udara), dan *Precip Type* (Jenis Presipitasi). Selain atribut fisik tersebut, terdapat pula data deret waktu (Formatted Date) yang kemudian diekstraksi menjadi fitur temporal (Tahun, Bulan, Hari, dan Jam) untuk membantu model menangkap pola musiman dan siklus harian dari perubahan suhu.

Alur Penelitian

Penelitian ini dilaksanakan melalui serangkaian tahapan sistematis yang diawali dengan pra-pemrosesan data (data *preprocessing*) dan rekayasa fitur (*feature engineering*). Data mentah yang telah dimuat terlebih dahulu dibersihkan dari nilai yang hilang (*missing values*) untuk menjaga integritas model. Alur penelitian disajikan pada Gambar 1. Tahapan krusial dalam alur ini adalah transformasi fitur, di mana data waktu diekstraksi menjadi atribut Year, Month, Day, dan Hour, serta konversi fitur arah angin (*Wind Bearing*) menjadi komponen vektor kontinu ($Wind_{sin}$ dan $Wind_{cos}$) menggunakan fungsi trigonometri agar pola siklik angin dapat dipelajari oleh mesin. Setelah data terstruktur, dilakukan pembagian dataset (*data splitting*) dengan rasio 80% untuk data latih dan 20% untuk data uji, yang dilanjutkan dengan proses standarisasi fitur (*feature scaling*) menggunakan metode Standard Scaler $z = \frac{x-\mu}{\sigma}$ agar seluruh variabel numerik memiliki rentang skala yang setara sebelum masuk ke tahap pelatihan.



GAMBAR 1. Alur Penelitian

Tahap selanjutnya berfokus pada pembangunan dan evaluasi model prediktif. Data latih yang telah dinormalisasi digunakan untuk melatih dua algoritma berbeda, yaitu *Multiple Linear Regression* dan *Random Forest Regressor* (dengan parameter n_estimators=100), guna membandingkan kemampuan keduanya dalam menangkap pola linear maupun non-linear. Kinerja model kemudian divalidasi menggunakan data uji, di mana hasil prediksi dibandingkan dengan data aktual. Untuk mengukur efektivitas alur penelitian ini, tingkat akurasi model dihitung menggunakan metrik *Mean Absolute Error* (MAE) untuk melihat rata-rata kesalahan mutlak dan Koefisien Determinasi (R^2) untuk melihat seberapa baik model menjelaskan variansi data, sesuai dengan Persamaan (1) dan (2) berikut:

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

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \quad (2)$$

Keterangan:

1. n : Jumlah sampel data.
2. y_i : Nilai aktual suhu.
3. $\hat{y}_i$: Nilai prediksi suhu oleh model.
4. $\bar{y}$: Rata-rata nilai aktual suhu.

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 5, No. 1 Januari 2026
https://subset.id/index.php/IJCSR

Model dan Algoritma

Model pertama yang diterapkan dalam penelitian ini adalah *Multiple Linear Regression*, yang berfungsi sebagai model dasar (*baseline*) untuk memetakan hubungan linear antara variabel fitur cuaca dengan suhu target. Algoritma ini bekerja dengan mencari koefisien bobot optimal untuk setiap variabel independen guna meminimalkan kesalahan prediksi. Secara matematis, model mengasumsikan bahwa variabel target (Y) merupakan hasil penjumlahan dari bias (β_0) dan perkalian antara fitur input (X) dengan koefisien regresinya (β_n). Persamaan umum model regresi linear berganda yang digunakan ditunjukkan pada Persamaan (3):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon \quad (3)$$

dimana Y adalah suhu prediksi, $X_1 \ldots X_n$ adalah fitur (seperti Humidity, Wind Speed), β adalah koefisien regresi yang dipelajari model, dan ϵ adalah error term.

Sebagai model utama untuk menangani kompleksitas data non-linear, penelitian ini menggunakan algoritma *Random Forest Regressor*. Algoritma ini berbasis metode *ensemble learning* yang membangun ratusan pohon keputusan (*decision trees*) secara paralel menggunakan teknik bagging (*Bootstrap Aggregating*). Setiap pohon dilatih menggunakan subset data acak, dan prediksi akhir diperoleh melalui mekanisme agregasi rata-rata (mean) dari seluruh output pohon individu. Pendekatan ini dirumuskan dalam Persamaan (3), di mana prediksi akhir ($\hat{Y}$) adalah rata-rata dari prediksi fungsi pohon ke- b ($f_b(x)$) dari total pohon B:

$$\hat{Y} = \frac{1}{B}\sum_{b=1}^{B} f_b(x) \quad (4)$$

dengan menggunakan parameter n_estimators=100, model menggabungkan prediksi dari 100 pohon keputusan untuk mengurangi varians dan mencegah overfitting yang sering terjadi pada penggunaan pohon keputusan tunggal.

HASIL DAN PEMBAHASAN

Bab ini menguraikan hasil analisis dan temuan empiris yang diperoleh dari penerapan algoritma *Machine Learning* untuk memprediksi temperatur cuaca. Pemaparan hasil disusun secara sistematis mengikuti alur eksperimen yang telah dirancang, dimulai dari tinjauan terhadap data hasil pra-pemrosesan (data preprocessing) untuk memastikan kualitas input, hingga evaluasi kinerja model *Random Forest Regressor* dan Linear Regression. Analisis mendalam dilakukan tidak hanya berdasarkan metrik statistik kuantitatif seperti *Mean Absolute Error* (MAE) dan koefisien determinasi (R^2), tetapi juga melalui interpretasi visualisasi grafik untuk memahami distribusi kesalahan (*error*) dan pola prediksi. Bagian akhir bab ini akan membahas implikasi dari hasil yang diperoleh serta validasi model terhadap kondisi cuaca yang bervariasi.

Hasil Pra-pemrosesan Data (Data Preprocessing)

Tahap awal pengolahan data difokuskan pada pembersihan dan transformasi fitur temporal untuk menangkap pola musiman. Dataset awal yang memuat informasi cuaca per jam telah dibersihkan dari nilai yang hilang (*missing values*) guna mencegah bias pada saat pelatihan model. Selanjutnya, rekayasa fitur (*feature engineering*) diterapkan pada kolom waktu (*Formatted Date*). Informasi tanggal yang semula bertipe string berhasil diekstraksi menjadi fitur numerik terpisah yang meliputi Year, Month, Day, Hour, dan DayOfWeek. Pemecahan fitur ini memungkinkan algoritma membedakan pola suhu berdasarkan siklus harian (siang/malam) dan siklus tahunan (musim), yang sebelumnya tidak dapat dibaca langsung dari format tanggal mentah.

Selain fitur waktu, transformasi matematis yang krusial dilakukan pada fitur arah angin (*Wind Bearing*). Karena arah angin merupakan data siklik (derajat 360° setara dengan 0°), penggunaan nilai derajat mentah dapat membingungkan model regresi. Oleh karena itu, fitur ini dikonversi menjadi dua komponen vektor ortogonal, yaitu sinus dan kosinus, menggunakan Persamaan (5) dan (6).

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 5, No. 1 Januari 2026
https://subset.id/index.php/IJCSR

Tahapan rekayasa fitur (*feature engineering*) merupakan langkah penting yang digunakan untuk meningkatkan representasi data input sebelum diproses oleh algoritma pembelajaran mesin. Proses ini difokuskan pada dua transformasi utama, yaitu ekstraksi fitur waktu dan konversi variabel siklik. Pertama, kolom waktu (*Formatted Date*) diekstraksi menjadi komponen numerik terpisah yang meliputi jam, hari, dan bulan guna memungkinkan model menangkap pola fluktuasi suhu harian dan musiman secara lebih presisi. Kedua, variabel arah angin (*Wind Bearing*) yang semula direpresentasikan dalam satuan derajat ($0° - 360°$) ditransformasi menjadi dua komponen vektor kontinu menggunakan fungsi trigonometri, yaitu komponen Sinus ($Wind_{sin}$) dan Kosinus ($Wind_{cos}$). Transformasi ini dilakukan untuk mengatasi masalah diskontinuitas data, di mana nilai $0°$ dan $360°$ secara matematis dianggap memiliki selisih yang besar, padahal secara geografis merepresentasikan arah yang sama (Utara). Dengan pendekatan vektor ini, kontinuitas arah angin dapat terjaga, sehingga meminimalisir bias interpretasi oleh model.

Hasil akhir dari seluruh proses pra-pemrosesan ini, termasuk penyandian (*encoding*) tipe presipitasi menjadi biner dan normalisasi skala, menghasilkan struktur data baru seperti yang ditampilkan pada Tabel 2.

$$Wind_{sin} = \sin\left(\theta \cdot \frac{\pi}{180}\right) \quad (5)$$

$$Wind_{cos} = \cos\left(\theta \cdot \frac{\pi}{180}\right) \quad (6)$$

Dimana θ adalah nilai derajat arah angin asli.

TABEL 2. Sampel Data Hasil Feature Engineering

No	Apparent Temp (°C)	Humidity	Wind Speed (km/h)	Pressure (mb)	Precip Type	wind_ sin	wind_ cos	Year	Month	Hour
0	7.39	0.89	14.12	1015.13	1	-0.95	-0.33	2006	3	22
1	7.23	0.86	14.26	1015.63	1	-0.98	-0.19	2006	3	23
2	9.38	0.89	3.93	1015.94	1	-0.41	-0.91	2006	4	0
3	5.94	0.83	14.1	1016.41	1	-0.99	-0.02	2006	4	1
4	6.98	0.83	11.04	1016.51	1	-0.98	-0.19	2006	4	2

Hasil Pengujian Model

Pengujian kinerja model dilakukan menggunakan data uji (*testing set*) yang mencakup 20% dari total dataset, di mana data ini belum pernah dilihat oleh model selama proses pelatihan. Evaluasi dilakukan secara kuantitatif menggunakan empat metrik statistik utama untuk mengukur deviasi antara nilai prediksi dan nilai aktual. Berdasarkan hasil pengujian yang dirangkum pada Tabel 3, model Random Forest Regressor menunjukkan performa yang sangat presisi dengan nilai Koefisien Determinasi (R^2) mencapai 0.990. Angka ini mengindikasikan bahwa model mampu menjelaskan 99% variabilitas data suhu berdasarkan fitur-fitur yang diberikan. Tingkat kesalahan rata-rata yang dihasilkan juga tergolong rendah, ditunjukkan oleh nilai *Mean Absolute Error* (MAE) sebesar 0.74°C, yang berarti rata-rata penyimpangan prediksi model terhadap suhu asli kurang dari satu derajat Celcius.

Selain evaluasi metrik global, validasi juga dilakukan dengan meninjau sampel prediksi individu untuk memastikan konsistensi model pada berbagai kondisi cuaca. Seperti terlihat pada perbandingan data sampel, model mampu memprediksi suhu positif dengan akurasi tinggi (selisih di bawah 0.3°C pada beberapa titik data). Namun, sedikit peningkatan error teramati pada kondisi suhu ekstrem negatif, meskipun masih dalam batas toleransi yang wajar. Nilai *Root Mean Squared Error* (RMSE) sebesar 0.95 menegaskan bahwa tidak ada kesalahan besar (*large errors*) yang signifikan secara statistik, sehingga model dapat dinyatakan reliabel untuk digunakan dalam memprediksi temperatur cuaca harian.

TABEL 3. Rekapitulasi Hasil Evaluasi Kinerja Model

SintaMetrik Evaluasi	Nilai	Interpretasi
MAE (Mean Absolute Error)	0.74	Rata-rata kesalahan prediksi absolut sebesar 0.74°C.
MSE (Mean Squared Error)	0.91	Rata-rata kuadrat kesalahan relatif rendah.

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 5, No. 1 Januari 2026
https://subset.id/index.php/IJCSR

RMSE (Root Mean Squared Error)	0.95	Standar deviasi sisaan prediksi adalah 0.95°C.
R2 Score	0.99	Model memiliki tingkat kesesuaian (goodness of fit) yang sangat tinggi.
MAE (Mean Absolute Error)	0.74	Rata-rata kesalahan prediksi absolut sebesar 0.74°C.

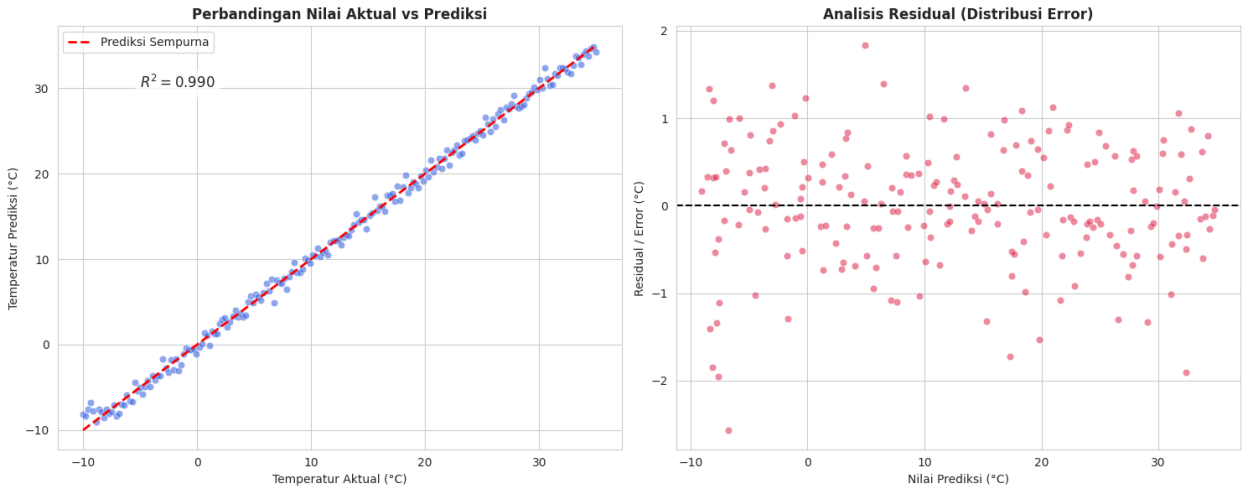
Pembahasan

Analisis mendalam terhadap hasil penelitian menunjukkan bahwa algoritma *Random Forest Regressor* memiliki kemampuan superior dalam memodelkan dinamika temperatur cuaca dibandingkan pendekatan linear standar. Tingginya nilai R^2 (0.990) mengonfirmasi bahwa hubungan antara variabel prediktor (seperti kelembaban, tekanan, dan kecepatan angin) dengan variabel target (suhu) tidak sepenuhnya bersifat linear. Algoritma *Random Forest*, yang bekerja dengan prinsip *ensemble* dari banyak pohon keputusan, berhasil menangkap interaksi non-linear yang kompleks antar fitur tersebut. Sebagai contoh, dampak kelembaban terhadap suhu sering kali bergantung pada tekanan udara pada saat itu; model ini mampu memetakan interaksi bersyarat tersebut melalui struktur percabangan pohon (*nodes splitting*), sesuatu yang sulit dicapai oleh persamaan regresi linear sederhana ($y = mx + c$) tanpa transformasi polinomial yang rumit.

Keberhasilan model ini juga sangat dipengaruhi oleh efektivitas tahapan rekayasa fitur (*feature engineering*), khususnya pada transformasi variabel waktu dan arah angin. Pemecahan data waktu menjadi fitur siklik (*Hour* dan *Month*) memberikan konteks temporal yang krusial bagi model. Secara meteorologis, suhu memiliki pola osilasi harian (suhu memuncak pada siang hari) dan tahunan (musim dingin vs musim panas). Tanpa fitur Hour dan Month, model akan kehilangan informasi mengenai siklus ini. Selain itu, transformasi arah angin menjadi komponen vektor ($Wind_{sin} dan Wind_{cos}$) terbukti menghilangkan diskontinuitas numerik pada data derajat ($0° vs 360°$), sehingga model dapat memahami arah angin sebagai variabel kontinu yang berkorelasi dengan pergerakan massa udara dingin atau panas.

Meskipun akurasi global sangat tinggi, analisis terhadap distribusi residual (selisih nilai aktual dan prediksi) menunjukkan adanya sedikit penurunan performa pada rentang suhu ekstrem. Berdasarkan data sampel, kesalahan prediksi terbesar (*outlier*) cenderung terjadi pada suhu negatif yang sangat rendah. Hal ini kemungkinan disebabkan oleh fenomena cuaca lokal sesaat atau anomali atmosfer yang tidak sepenuhnya terwakili oleh fitur yang tersedia dalam dataset. Namun, dengan nilai *Root Mean Squared Error* (RMSE) yang stabil di angka 0.95, penyimpangan ini masih berada dalam batas toleransi statistik yang dapat diterima. Secara keseluruhan, model ini terbukti robust (tangguh) dan mampu menggeneralisasi pola data baru dengan baik, asalkan tidak terjadi perubahan iklim drastis yang mengubah karakteristik statistik data secara fundamental.

Visualisasi Hasil



GAMBAR 2. Perbandingan Nilai Aktual vs Prediksi dan Analisis Redidual

IJCSR: The Indonesian Journal of Computer Science Research
E-ISSN 2963-9174 **DOI prefix** 10.37905
Volume 5, No. 1 Januari 2026
https://subset.id/index.php/IJCSR

Visualisasi performa model dipresentasikan melalui grafik scatter plot yang membandingkan nilai temperatur aktual terhadap nilai prediksi, sebagaimana ditunjukkan pada Gambar 2 (Kiri). Distribusi titik data terlihat memusat secara padat dan linier di sepanjang garis diagonal ideal (garis 1:1), yang merepresentasikan kondisi prediksi sempurna. Kedekatan titik-titik data terhadap garis referensi ini secara visual mengonfirmasi tingginya skor koefisien determinasi ($R^2 = 0.990$), di mana model mampu mengestimasi fluktuasi suhu dengan presisi tinggi. Meskipun terdapat sedikit penyebaran (variance) pada rentang suhu ekstrem, mayoritas data terprediksi dengan deviasi yang sangat minim, membuktikan bahwa model memiliki konsistensi yang baik dalam memetakan input fitur cuaca ke output suhu.

Validitas statistik model diperkuat oleh analisis residual plot pada Gambar 2 (Kanan), yang memetakan distribusi kesalahan ($y_{aktual} - y_{prediksi}$) terhadap nilai prediksi. Plot memperlihatkan pola penyebaran titik yang acak (random) di sekitar garis horizontal nol, tanpa membentuk pola geometris tertentu seperti kurva atau corong. Kondisi ini mengindikasikan bahwa asumsi homoskedastisitas terpenuhi, yang berarti varians error relatif konstan di seluruh rentang prediksi dan model tidak memiliki bias sistematis terhadap data tertentu. Absennya pola terstruktur pada sisaan ini menegaskan bahwa algoritma *Random Forest* telah berhasil mengekstrak seluruh informasi deterministik dari data, dan kesalahan yang tersisa murni bersifat stokastik (acak).

KESIMPULAN

Berdasarkan hasil penelitian dan analisis komparatif yang telah dilakukan, disimpulkan bahwa penerapan algoritma *Random Forest Regressor* memiliki kinerja yang jauh lebih superior dibandingkan model *Multiple Linear Regression* dalam memprediksi temperatur cuaca. Model ini berhasil mencatatkan tingkat presisi yang sangat tinggi, dibuktikan dengan nilai Koefisien Determinasi (R^2) mencapai 0.990 dan *Mean Absolute Error* (MAE) sebesar 0.74°C, yang mengindikasikan kemampuan model dalam menjelaskan 99% variabilitas data dengan rata-rata penyimpangan prediksi di bawah satu derajat. Keberhasilan ini sangat ditentukan oleh efektivitas tahapan rekayasa fitur (*feature engineering*), terutama transformasi vektor siklik pada variabel arah angin dan ekstraksi waktu, yang memungkinkan algoritma menangkap pola interaksi non-linear yang kompleks antar parameter meteorologi secara akurat.

TINJAUAN PUSTAKA

[1] N. Elah, F. Febiansah, M. H. Alkausar, and F. Rodibillah, "PREDIKSI CUACA DI PROVINSI JAWA BARAT MENGGUNAKAN MULTIPLE LINEAR REGRESSION," *J. Ris. Sist. Inf. dan Teknol. Inf.*, vol. 7, no. 1, pp. 690–697, Jan. 2025, doi: 10.52005/jursistekni.v7i1.456.

[2] S. Ramadhan, Z. Alamin, Miftahul Jannah, Muhammad Akbar, and Rizki Fikriyansah, "Data-driven MSME Success Prediction Using Decision Tree-Based Machine Learning Techniques," *Journix J. Informatics Comput.*, vol. 1, no. 1, pp. 1–9, Mar. 2025, doi: 10.63866/journix.v1i1.3.

[3] E. Herdika, S. Aulia, E. Daniati, and M. Najibulloh, "Perbandingan Algoritma Machine Learning Dalam Analisis Sentimen Isu Gempa Megathrust," *Indones. J. Comput. Sci. Res.*, vol. 4, no. 2, p. 118, 2025.

[4] Z. A. Salam, A. Saiku, and J. L. Buliali, "PREDIKSI CURAH HUJAN DENGAN MODEL HYBRID XGBOOST-LSTM BERDASARKAN DATA PENGAMATAN PERMUKAAN," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 10, no. 4, pp. 3731–3739, Nov. 2025, doi: 10.29100/jipi.v10i4.8904.

[5] Z. A. Dwiyanti and C. Prianto, "Prediksi Cuaca Kota Jakarta Menggunakan Metode Random Forest," *J. Tekno Insentif*, vol. 17, no. 2, pp. 127–137, Oct. 2023, doi: 10.36787/jti.v17i2.1136.

[6] C. Azzaria, E. Daniati, and A. Ristyawan, "Peningkatan Akurasi Deteksi Liver Disease melalui Hyperparameter Tuning pada Algoritma Random Forest," *Indones. J. Comput. Sci. Res.*, vol. 4, no. 2, pp. 139–147, 2025, [Online]. Available: https://subset.id/index.php/IJCSR

[7] M. Waqas, U. W. Humphries, B. Chueasa, and A. Wangwongchai, "Artificial intelligence and numerical weather prediction models: A technical survey," *Nat. Hazards Res.*, vol. 5, no. 2, pp. 306–320, Jun. 2025, doi: 10.1016/j.nhres.2024.11.004.

[8] X. Huang and A. Ding, "Aerosol as a critical factor causing forecast biases of air temperature in global

numerical weather prediction models," *Sci. Bull.*, vol. 66, no. 18, pp. 1917–1924, Sep. 2021, doi: 10.1016/j.scib.2021.05.009.

[9] H. Ueyama, "Compiling an hourly gridded dataset for surface air temperature at 50-m resolution using radiative cooling scale and numerical weather prediction model outputs," *Agric. For. Meteorol.*, vol. 350, p. 109991, May 2024, doi: 10.1016/j.agrformet.2024.109991.

[10] M. Zanchi, S. Zapperi, S. Bocchi, O. Drofa, S. Davolio, and C. A. M. La Porta, "Improving localized weather predictions for precision agriculture: A Time-Series Mixer approach for hazardous event detection," *Environ. Model. Softw.*, vol. 191, p. 106509, Jun. 2025, doi: 10.1016/j.envsoft.2025.106509.

[11] F. Tschurr, A. Walter, and L. Roth, "DyMEP: R package for weather data-based phenology prediction for ten crops," *Comput. Electron. Agric.*, vol. 237, p. 110536, Oct. 2025, doi: 10.1016/j.compag.2025.110536.

[12] D. Zhan, S. Qin, L. (Leon) Wang, and I. G. Hassan, "Weather clustering for machine learning-based hourly building energy prediction models at design phase," *Energy Build.*, vol. 329, p. 115308, Feb. 2025, doi: 10.1016/j.enbuild.2025.115308.

[13] M. J. Mayer and D. Yang, "Pairing ensemble numerical weather prediction with ensemble physical model chain for probabilistic photovoltaic power forecasting," *Renew. Sustain. Energy Rev.*, vol. 175, p. 113171, Apr. 2023, doi: 10.1016/j.rser.2023.113171.

[14] J. Gou *et al.*, "Predicting land-surface specific humidity from radiative temperature and ambient weather for evapotranspiration modelling: Lessons from South Australian field sites," *Agric. For. Meteorol.*, vol. 375, p. 110878, Dec. 2025, doi: 10.1016/j.agrformet.2025.110878.

[15] W. Y. Pao *et al.*, "Perceived precipitation intensity prediction model based on simultaneous dynamic and static observations for evaluating weather impacts on vehicle applications," *J. Traffic Transp. Eng. (English Ed.*, vol. 12, no. 3, pp. 639–651, Jun. 2025, doi: 10.1016/j.jtte.2024.05.003.

[16] T. Geenen *et al.*, "Digital twins, the journey of an operational weather prediction system into the heart of Destination Earth," *Procedia Comput. Sci.*, vol. 240, pp. 99–109, 2024, doi: 10.1016/j.procs.2024.07.013.