

Prediksi Data Cuaca Menggunakan A.I. *Random Forest Regressor* pada Kota Szeged, Hungaria

Darrel William¹, I Wayan Sudiarsa²,
Varren Williams³, Andre Sifani Putra⁴, Zacky Ade Arjana⁵
^{1,2,3,4,5} Institut Bisnis dan Teknologi Indonesia
E-mail: sudiarsa@instiki.ac.id

Article Info

Article history:

Received January 02, 2025
Revised January 10, 2025
Accepted January 18, 2025

Keywords:

Random Forest Regressor,
Temperature Prediction,
Machine Learning, Meteorology.

ABSTRACT

This research aims to predict weather data in Szeged, Hungary, using the Machine Learning algorithm Random Forest Regressor. The data used covers a ten-year period (04/01/2006 – 09/09/2016) with input parameters including wind direction, visibility, humidity, wind speed, and air pressure to predict the temperature variable. The research methodology includes data pre-processing, dataset splitting (80% training data, 20% testing data), and model evaluation using statistical metrics. The research results show that the Random Forest model is able to capture seasonal trends with a Mean Absolute Error (MAE) value of 4.882, a Root Mean Squared Error (RMSE) value of 6.102, and a coefficient of determination (R^2) value of 0.560. Although the model showed fairly stable performance in predicting temperature trends during the 2015-2016 period, the R^2 value indicates that there is still room for further optimization through hyperparameter adjustments or the addition of other supporting features. This research contributes as an initial reference in the application of artificial intelligence for meteorological analysis in an academic setting.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Article Info

Article history:

Received January 02, 2025
Revised January 10, 2025
Accepted January 18, 2025

Keywords:

Random Forest Regressor,
Prediksi Suhu, Machine
Learning, Data Cuaca,
Meteorologi

ABSTRAK

Penelitian ini bertujuan untuk memprediksi data cuaca di Kota Szeged, Hungaria, dengan memanfaatkan algoritma Machine Learning yaitu Random Forest Regressor. Data yang digunakan mencakup periode sepuluh tahun (01/04/2006 – 09/09/2016) dengan parameter input berupa arah angin, jarak pandang, humiditas, kecepatan angin, dan tekanan udara untuk memprediksi variabel temperatur. Metodologi penelitian meliputi pra-pemrosesan data, pembagian dataset (80% latih, 20% uji), dan evaluasi model menggunakan metrik statistik. Hasil penelitian menunjukkan bahwa model Random Forest mampu menangkap tren musiman dengan nilai Mean Absolute Error (MAE) sebesar 4,882, Root Mean Squared Error (RMSE) sebesar 6,102, dan koefisien determinasi (R^2) sebesar 0,560. Meskipun model menunjukkan performa yang cukup stabil dalam memprediksi tren suhu pada periode 2015-2016, nilai R^2 menunjukkan bahwa masih terdapat ruang untuk optimasi lebih lanjut melalui penyesuaian hyperparameter atau penambahan fitur pendukung lainnya. Penelitian ini memberikan kontribusi sebagai referensi awal dalam penerapan kecerdasan buatan untuk analisis meteorologi di lingkungan akademik.



Corresponding Author:

I Wayan Sudiarsa

Institut Bisnis dan Teknologi Indonesia

Email: sudiarsa@instiki.ac.id

Pendahuluan

Cuaca merupakan fenomena meteorologis yang kompleks dan dinamis yang memengaruhi banyak sektor kehidupan manusia, termasuk pertanian, transportasi, kesehatan, dan perencanaan kota. Ketidakpastian kondisi cuaca sering kali membawa dampak signifikan, sehingga prediksi cuaca yang akurat menjadi kebutuhan penting untuk mitigasi risiko serta pengambilan keputusan yang efektif dalam kegiatan sosial dan ekonomi. Dengan meningkatnya volume data cuaca historis yang tersedia, pendekatan machine learning menjadi pilihan yang populer dalam upaya meningkatkan akurasi prediksi dibandingkan dengan metode konvensional berbasis persamaan fisika saja.

Dalam beberapa tahun terakhir, metode machine learning telah banyak diterapkan untuk prediksi cuaca melalui analisis pola pada data historis. Salah satu algoritma yang menonjol dalam tugas regresi adalah Random Forest Regressor, yang bekerja berdasarkan prinsip ensemble learning dengan membangun banyak pohon keputusan (decision tree) untuk menghasilkan prediksi akhir yang lebih stabil dan akurat. Penerapan Random Forest dalam studi prediksi cuaca menunjukkan kinerja yang baik dibandingkan dengan algoritma lain, dengan kemampuan menangani variabilitas data yang kompleks dan meminimalkan risiko overfitting pada model (Comprehensive Approach to Weather Prediction with the Random Forest Algorithm, 2025).

Selain itu, penelitian-penelitian terbaru juga menegaskan bahwa pemilihan fitur (feature selection) dan penggunaan algoritma Random Forest dapat berdampak positif terhadap akurasi prediksi cuaca. Studi yang mengevaluasi pengaruh pemilihan fitur menemukan bahwa beberapa variabel cuaca memiliki kontribusi penting terhadap kinerja model, sehingga pengolahan data yang tepat menjadi aspek penting dalam pengembangan sistem prediktif (Weather Prediction with Feature Selection and Random Forest, 2024).

Potensi penggunaan machine learning dalam prakiraan cuaca juga didukung oleh penelitian empiris di berbagai konteks geografis. Misalnya, dalam studi prediksi cuaca di suatu wilayah, algoritma Random Forest mampu memberikan estimasi kejadian cuaca yang efektif setelah data dipraproses dan divisualisasikan secara terstruktur, sehingga memperkuat posisi metode ini sebagai salah satu pendekatan yang layak dalam pengembangan sistem prediksi berbasis data (Analisis Model Prediksi Cuaca Menggunakan Support Vector Machine, Gradient Boosting, Random Forest, dan Decision Tree, 2025).

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis prediksi kondisi cuaca pada Kota Leeds, England menggunakan metode Random Forest Regressor. Penelitian ini diharapkan dapat memberikan gambaran lebih lanjut mengenai performa model

dalam memprediksi parameter cuaca serta memberikan kontribusi terhadap literatur ilmiah dalam bidang prediksi cuaca berbasis machine learning

Metode

Penelitian ini menerapkan pendekatan kuantitatif eksperimental untuk memprediksi kondisi cuaca memanfaatkan algoritma *Random Forest Regressor*. Dataset yang digunakan merupakan rekaman cuaca historis dari kota Szeged, Hungaria, yang mencakup variabel seperti suhu, kelembapan, tekanan udara, dan kecepatan angin (Wilks, 2011).

Proses penelitian diawali dengan tahap pra-pemrosesan data, yang meliputi penanganan *missing values*, seleksi fitur relevan, serta normalisasi data guna menyamakan skala antar variabel. Langkah ini krusial untuk mengoptimalkan kinerja model *machine learning* (Hastie et al., 2009). Selanjutnya, dataset dipartisi dengan rasio 80% sebagai data latih (*training data*) untuk pembentukan model, dan 20% sebagai data uji (*testing data*) untuk validasi performa pada data baru (Géron, 2019).

Algoritma *Random Forest Regressor* bekerja dengan mengkonstruksi sejumlah pohon keputusan (*decision trees*) secara acak, kemudian mengagregasi hasil prediksinya untuk mendapatkan nilai akhir. Metode ini dipilih karena keandalannya dalam menangani data non-linier dan meminimalisir risiko *overfitting* (Breiman, 2001). Evaluasi akurasi model dilakukan menggunakan metrik *Mean Absolute Error* (MAE) dan *Root Mean Squared Error* (RMSE), di mana nilai kesalahan yang rendah mengindikasikan kualitas peramalan yang lebih presisi (Chai & Draxler, 2014).

Hasil & Pembahasan

a. Pra Pemrosesan Data

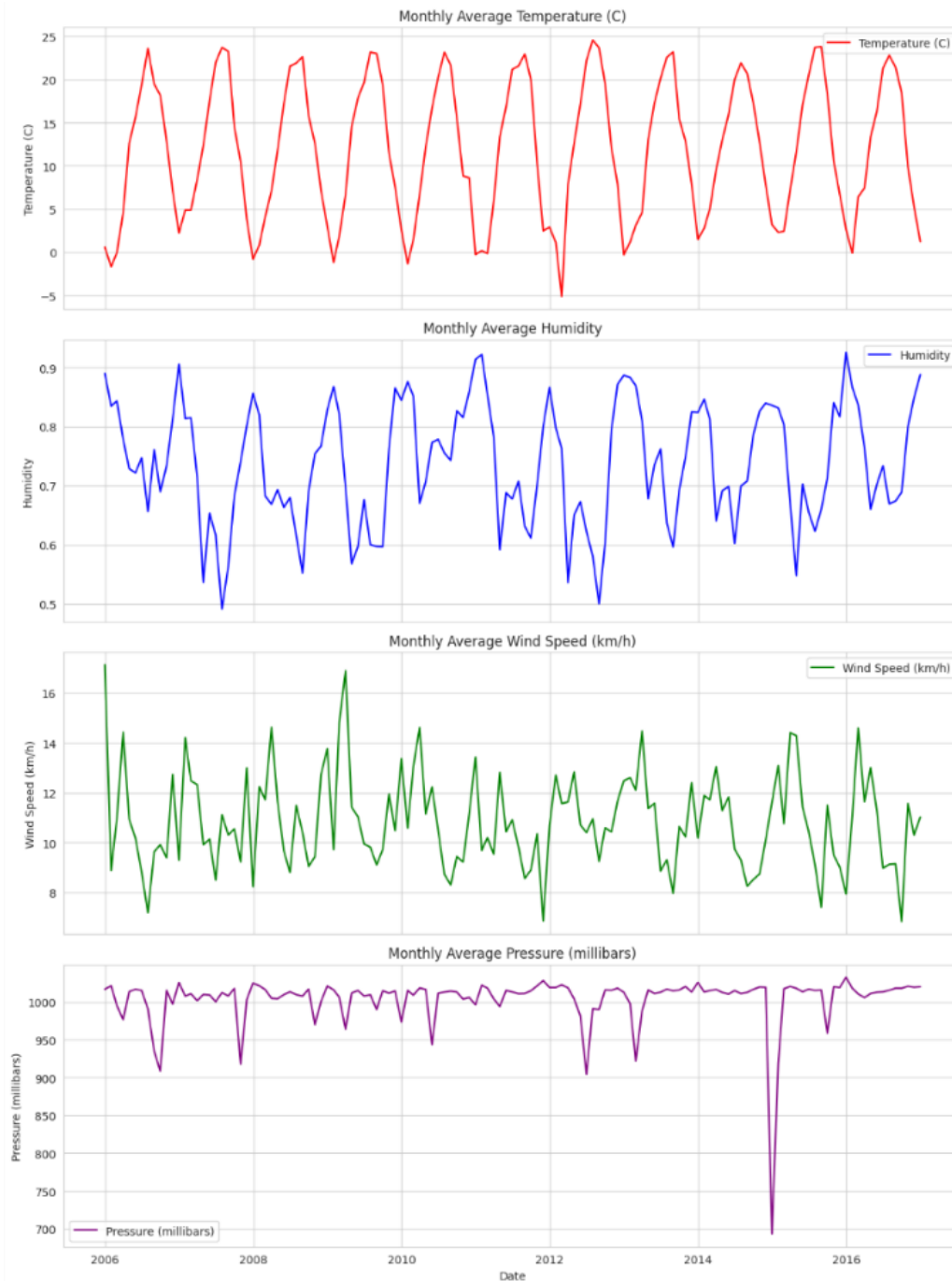
Pra-pemrosesan data dimulai dari mendeteksi dan membersihkan data-data yang bersifat tidak konklusif ringkasan dari dataset yang telah teranalisa kami tampilkan pada tabel berikut.

Tabel Ringkasan Data Inkonklusif

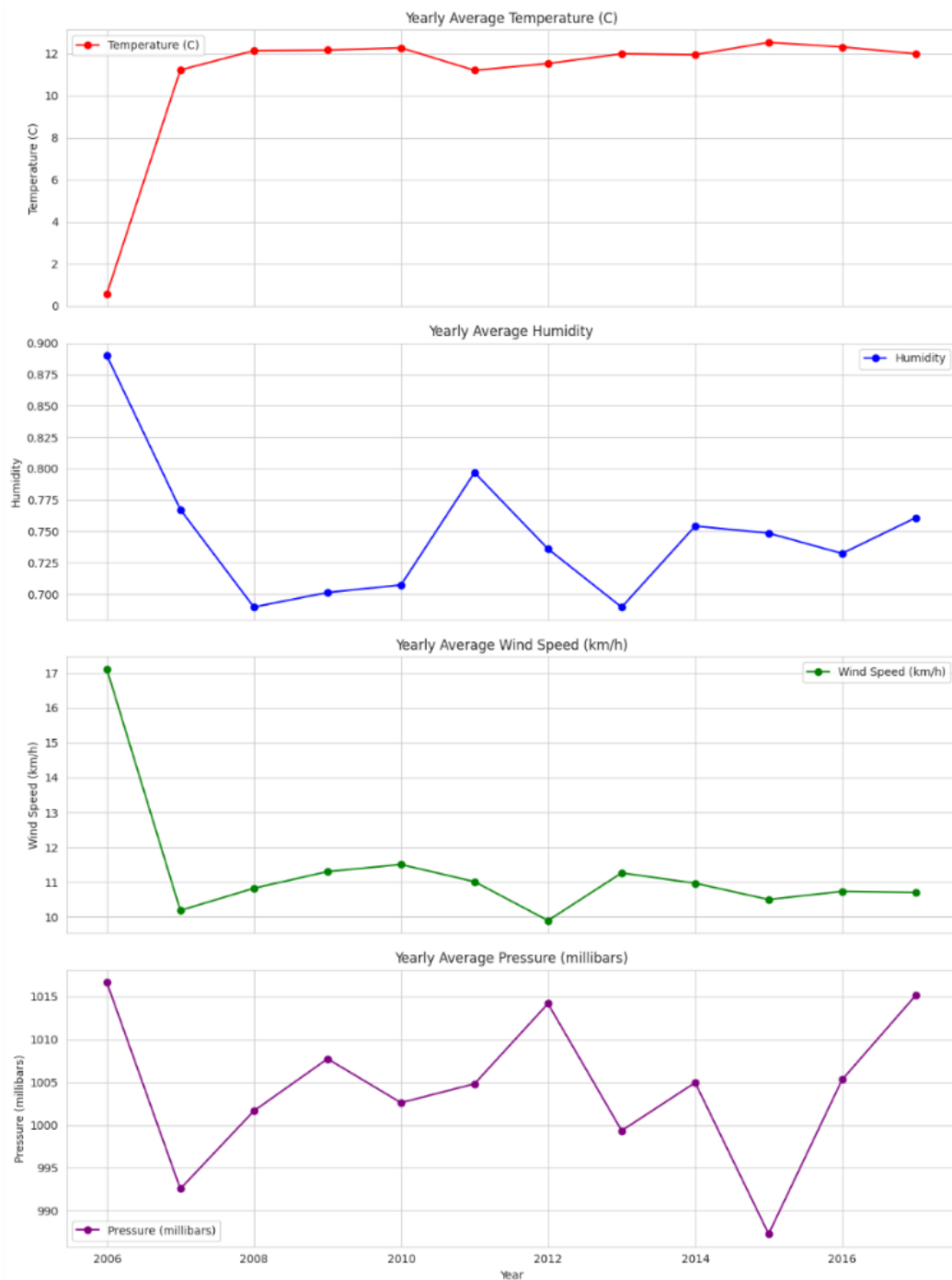
Jenis Masalah	Jumlah	Deskripsi
Baris Duplikat	24	Baris yang memiliki data identik sepenuhnya (data berulang).
Tipe Presipitasi Kosong	517	Kolom Precip Type tidak memiliki data (kosong/\$NaN\$).
Tekanan Udara Nol	1.288	Nilai Pressure (millibars) tercatat \$0\$, yang secara fisik tidak mungkin terjadi di permukaan bumi.
Kelembapan Nol	22	Pembacaan Humidity sebesar \$0\$, yang sangat tidak lazim untuk data cuaca terbuka.
Data Kosong (Loud Cover)	96.453	Seluruh baris pada kolom Loud Cover bernilai \$0\$,

sehingga tidak memberikan informasi apa pun.

b. Analisis Tren Cuaca; Rata-Rata per bulan dan per tahun



Figur 3.1

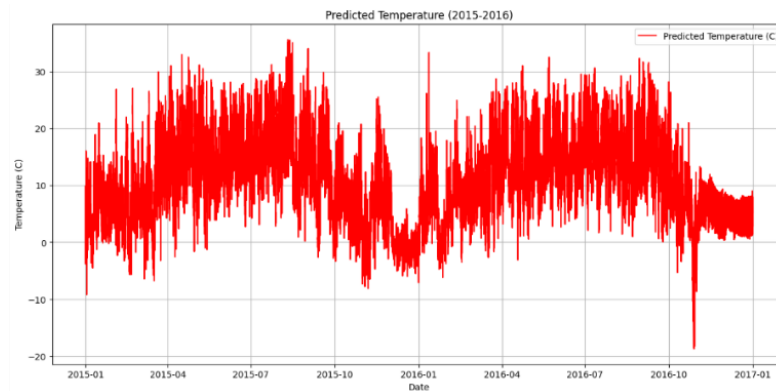


Figur 3.2

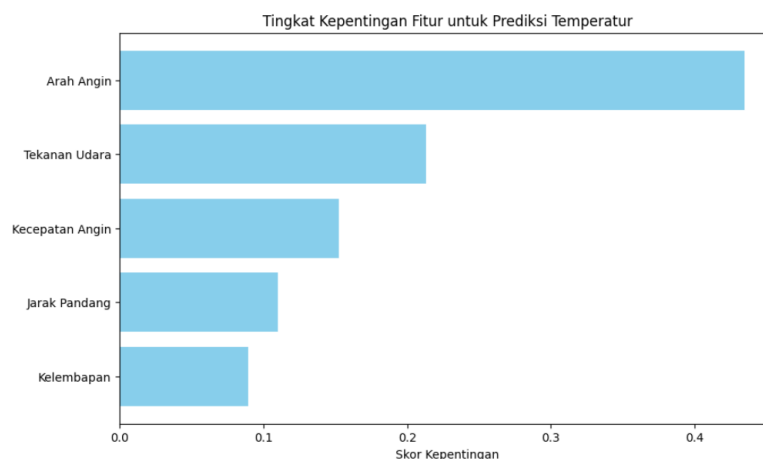
Kami gunakan dataset pasca pemrosesan untuk menampilkan sebuah visualisasi tren cuaca dengan parameter Temperatur, Humiditas, Kecepatan Angin, dan Tekanan, dimana variable rata-rata pada setiap parameter digunakan untuk menghasilkan tren parameter per bulan dan juga per tahun, grafik tersebut dapat dilihat pada *Figur 3.1* dan *Figur 3.2*

Hasil dari visualisasi ini bertujuan untuk menggambarkan lebih jelas tren cuaca yang terjadi pada Szeged, Hungaria, dimana data dari penjumlahan rata-rata parameter ini juga digunakan sebagai data *Machine Learning*.

- c. Machine Learning menggunakan data Analisis tren tergabung dengan data pasca pemrosesan awal untuk menghasilkan sebuah prediksi.



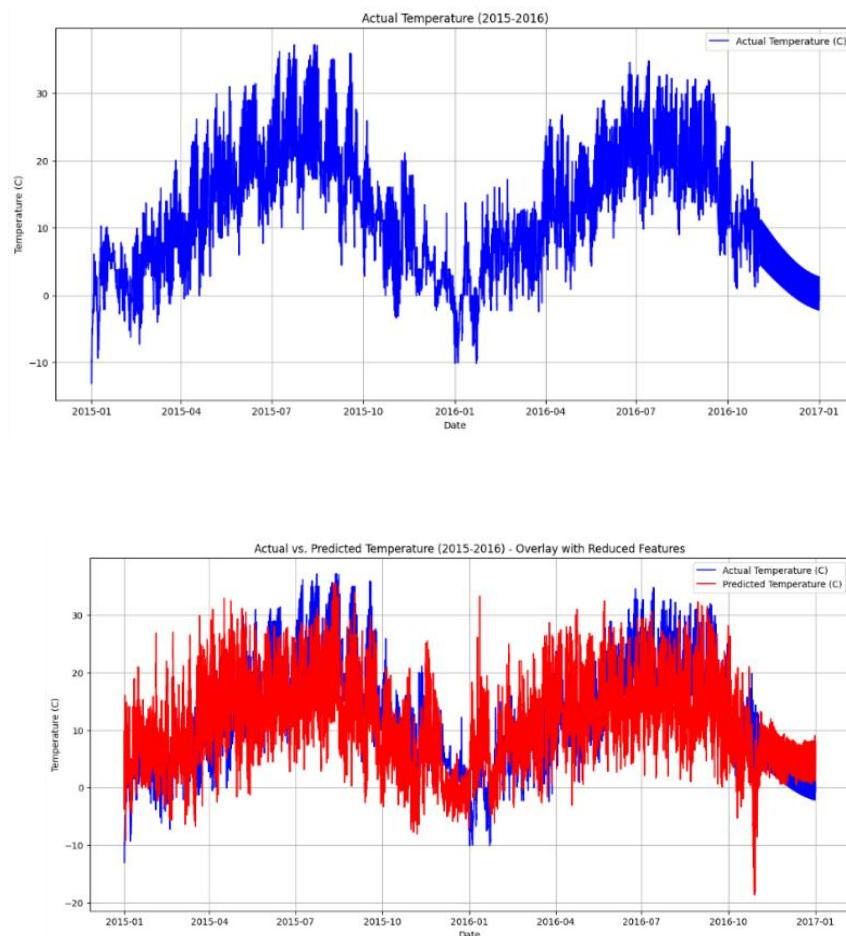
Figur 3.3



Figur 3.4

Ternyata parameter yang paling diandalkan oleh mesin adalah arah angin dengan jumlah Skor Kepentingan 0.4 yang dapat dilihat pada *Figur 3.4*, dimana parameter lain hanya berada di 0.22 untuk Tekanan Udara, 0.15 untuk Kecepatan Angin, 0.115 untuk Jarak Pandang, dan 0.09 untuk Kelembapan. Ternyata parameter yang paling diandalkan oleh mesin adalah arah angin dengan jumlah Skor Kepentingan 0.4 yang dapat dilihat pada *Figur 3.4*, dimana parameter lain hanya berada di 0.22 untuk Tekanan Udara, 0.15 untuk Kecepatan Angin, 0.115 untuk Jarak Pandang, dan 0.09 untuk Kelembapan. Hasil dari prediksi Temperatur dapat dilihat pada grafik di *Figur 3.3*

d. Perbandingan Data Prediksi dengan Data Nyata



Hasil Dari prediksi akhir dapat kami bandingkan secara langsung dengan data nyata dimana dengan analisa lebih lanjut menghasilkan variabel dengan kinerja pada tabel berikut.

Tabel Kinerja Perbandingan Data Prediksi Temperatur dengan Temperatur Asli

Mean Absolute Error (MAE):	4.882
Mean Squared Error (MSE):	37.239
Root Mean Squared Error (RMSE):	6.102
R-Squared (R^2):	0.560

Dimana pada tabel tersebut Mean Absolute Error (MAE) menyatakan jumlah variable meleset rata-rata sebesar 4.882, atau dalam konteks permasalahan ini berarti 4.882 Derajat Celsius. Mean Squared Error (MSE) adalah sebuah kuadrat yang didapatkan dari setiap selisih error dan kemudian diratakan, dikarenakan sifat kuadrat, ini berarti sangat sensitif terhadap

setiap error, dimana semakin besar nilai Mean Squared Error berarti menandakan semakin banyak kesalahan besar.

Root Mean Squared Error (RMSE) menyatakan besaran rata-rata kesalahan dalam satuan yang sama dengan variabel target. Karena dapat diamati bahwa nilai RMSE lebih besar daripada MAE itu berarti sesekali nilai dari model prediksi melakukan kesalahan yang cukup besar R-Squared (R^2) menyatakan model yang mampu menjelaskan sebesar 56% dari variasi data target, dimana 44% sisa yang tidak dapat dijelaskan kemungkinan karena faktor-faktor lain yang tidak termasuk pada model atau dari gangguan (noise).

Kesimpulan

Ringkasan Penelitian ini menganalisis data cuaca historis dari Szeged, Hungaria, yang mencakup periode 2006 hingga 2016. Menggunakan kumpulan data yang terdiri dari variabel seperti arah angin, jarak pandang, kelembapan, kecepatan angin, dan tekanan atmosfer, studi ini menerapkan model Random Forest Regressor untuk memprediksi fluktuasi suhu. Metodologi yang digunakan meliputi pra-pemrosesan data (pembersihan dan normalisasi), pembagian data 80/20 untuk pelatihan dan pengujian, dan evaluasi akurasi prediksi model terhadap data suhu aktual tahun 2015–2016. Hasilnya menunjukkan Mean Absolute Error (MAE) sebesar 4.882 dan nilai R-squared (R^2) sebesar 0.560, yang mengindikasikan korelasi sedang antara tren suhu yang diprediksi dan aktual tetapi dalam kenyataan kesalahan rata-rata sebesar 4.8 derajat celsius dalam kasus-kasus tertentu dapat dianggap perbedaan yang terlalu besar. Penerapan Random Forest Regressor dalam penelitian ini menunjukkan potensi pembelajaran ensemble untuk peramalan cuaca di wilayah Szeged. Meskipun model berhasil menangkap tren musiman umum dari data suhu, metrik kesalahan (RMSE sebesar 6.102) dan nilai R^2 menunjukkan bahwa akurasi model dapat ditingkatkan lebih lanjut. Kinerja prediksi saat ini menunjukkan bahwa meskipun fitur yang dipilih (kelembaban, tekanan, dll.) berpengaruh, pola cuaca di Szeged mungkin juga dipengaruhi oleh variabel non-linear yang tidak sepenuhnya ditangkap dalam konfigurasi model saat ini. Penelitian selanjutnya harus fokus pada pengoptimalan hyperparameter dalam algoritma Random Forest atau mengintegrasikan fitur tambahan, seperti indikator musiman atau data geografis tetangga, untuk meningkatkan ketelitian hasil peramalan. Secara keseluruhan, penelitian ini berfungsi sebagai langkah dasar untuk mengimplementasikan analisis meteorologi berbasis AI di Instiki.

Daftar Pustaka

- Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25(2), 197–227. <https://doi.org/10.1007/s11749-016-0481-7>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7(3), 1247–1250. <https://doi.org/10.5194/gmd-7-1247-2014>

- Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (2nd ed.). O'Reilly Media.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
- Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (3rd ed.). Academic Press.