



Analisis Performa Model *Random Forest* dan *Support Vector Regression* untuk Prediksi Suhu Maksimum Harian di Kota Jambi

R. Zaevan Khazafi Putra¹, Riza Pahlevi^{2*}, Ronald Naibaho³, Agus Nugroho⁴

^{1,2*}Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dinamika Bangsa, Jambi, Indonesia

³Fakultas Ilmu Komputer, Program Studi Sistem Informasi, Universitas Dinamika Bangsa, Jambi, Indonesia

⁴Fakultas Ilmu Komputer, Program Studi Sistem Komputer, Universitas Dinamika Bangsa, Jambi, Indonesia

Email: r.zaevan21@gmail.com¹, rizapahlevi@unama.ac.id^{2*}, rhodes8083@yahoo.co.id³, agusnugroho0888@gmail.com⁴

Alamat: Jl. Jend. Sudirman, The Hok, Kec. Jambi Selatan, Kota Jambi, Jambi 36138

*Penulis Korespondensi: rizapahlevi@unama.ac.id

Abstract. *The dynamic changes in weather patterns in Jambi City require an accurate temperature prediction system, thus this study aims to compare the performance of Random Forest and Support Vector Regression (SVR) algorithms in predicting daily maximum temperatures using weather data from 2020–2024 obtained from Open-Meteo with the application of Feature Engineering including lag and rolling window features. The test results indicate that the SVR model with a Radial Basis Function (RBF) kernel optimized using Grid Search ($C=10$, $\epsilon=0.2$, $\gamma=0.01$) significantly outperforms Random Forest based on a statistical Paired T-test (p -value < 0.05), yielding an R -squared (R^2) value of 87.46%, Mean Absolute Error (MAE) of 0.3818 °C, and Root Mean Squared Error (RMSE) of 0.4964 °C compared to Random Forest's R^2 of 84.05%, where the previous day's temperature (lag) and three-day rolling average were identified as the most dominant predictors, leading to the recommendation of SVR as the more effective method for temperature prediction in the study area.*

Keywords: *Feature Engineering, Maximum Temperature, Random Forest, Support Vector Regression, Weather Prediction.*

Abstrak. Perubahan pola cuaca yang dinamis di Kota Jambi menuntut adanya sistem prediksi suhu yang akurat, sehingga penelitian ini bertujuan untuk membandingkan kinerja algoritma *Random Forest* dan *Support Vector Regression* (SVR) dalam memprediksi suhu maksimum harian menggunakan data cuaca periode 2020–2024 dari *Open-Meteo* dengan penerapan *Feature Engineering* berupa fitur *lag* dan *rolling window*. Hasil pengujian menunjukkan bahwa model SVR dengan kernel *Radial Basis Function* (RBF) yang dioptimalkan menggunakan *Grid Search* ($C=10$, $\epsilon=0.2$, $\gamma=0.01$) terbukti lebih unggul secara signifikan berdasarkan uji statistik *Paired T-test* (p -value $< 0,05$), menghasilkan nilai R -squared (R^2) sebesar 87,46%, *Mean Absolute Error* (MAE) 0,3818 °C, dan *Root Mean Squared Error* (RMSE) 0,4964 °C, dibandingkan *Random Forest* yang hanya mencapai R^2 84,05%, di mana fitur suhu hari sebelumnya (*lag*) dan rata-rata tiga hari terakhir (*rolling*) teridentifikasi sebagai prediktor paling dominan, sehingga SVR direkomendasikan sebagai metode yang lebih efektif untuk prediksi suhu di wilayah studi.

Kata kunci: *Feature Engineering, Prediksi Cuaca, Random Forest, Suhu Maksimum, Support Vector Regression.*

1. LATAR BELAKANG

Prediksi cuaca yang akurat, khususnya suhu maksimum harian, memegang peranan vital dalam mendukung pengambilan keputusan strategis di berbagai sektor, mulai dari pertanian, manajemen energi, hingga mitigasi bencana (Gupta et al., 2022). Di Indonesia, tantangan prediksi cuaca semakin kompleks akibat karakteristik iklim tropis yang memiliki variabilitas tinggi dan sering kali dipengaruhi oleh fenomena anomali cuaca lokal. Metode konvensional

seperti *Numerical Weather Prediction* (NWP) yang berbasis pada pemodelan fisika atmosfer sering kali menghadapi kendala dalam hal resolusi spasial dan kebutuhan komputasi yang besar (Kumar & others, 2021), sehingga terkadang kurang responsif terhadap dinamika perubahan cuaca pada skala lokal yang spesifik seperti di Kota Jambi. Oleh karena itu, pendekatan berbasis data (*data-driven*) menggunakan algoritma *Machine Learning* (ML) telah muncul sebagai solusi alternatif yang menjanjikan, karena kemampuannya untuk memodelkan hubungan non-linear yang kompleks dari data historis tanpa bergantung pada asumsi fisik yang kaku (Béland & Sun, 2021; Xu & others, 2022).

Berbagai penelitian terdahulu telah mengeksplorasi penerapan algoritma ML dalam meteorologi dengan hasil yang bervariasi. Algoritma *Random Forest* (RF) sering menjadi pilihan utama karena ketahanannya terhadap *overfitting* dan kemampuannya menangani data dengan dimensi tinggi serta *noise* (Breiman, 2001), seperti yang ditunjukkan dalam studi prediksi curah hujan dan suhu di beberapa wilayah tropis (Indrasari & Suhendar, n.d.). Di sisi lain, *Support Vector Regression* (SVR) juga mendapatkan perhatian luas karena efektivitasnya dalam menangani dataset berukuran kecil hingga menengah dan kemampuannya melakukan generalisasi yang baik melalui penggunaan fungsi *kernel* non-linear (Smaili et al., 2022; Smola & Schölkopf, 2004). Namun, mayoritas studi yang ada cenderung berfokus pada dataset global atau wilayah subtropis, dan masih terbatasnya penelitian yang secara spesifik membandingkan efektivitas kedua algoritma ini (RF dan SVR) dalam konteks data iklim lokal Kota Jambi yang memiliki karakteristik fluktuasi suhu dan kelembapan yang unik.

Kesenjangan penelitian (*gap analysis*) utama yang diangkat dalam studi ini adalah belum adanya evaluasi komparatif yang mendalam mengenai kinerja RF dan SVR yang dioptimalkan khusus untuk prediksi suhu maksimum harian di Kota Jambi dengan memanfaatkan teknik rekayasa fitur (*feature engineering*) temporal. Kebaruan penelitian ini terletak pada integrasi fitur *lag* dan *rolling window* untuk menangkap pola autokorelasi dalam data deret waktu (Brownlee, 2020), serta penerapan optimasi *hyperparameter* yang ketat untuk memaksimalkan potensi masing-masing model. Penelitian ini bertujuan untuk mengisi kekosongan tersebut dengan mengembangkan, melatih, dan memvalidasi model prediksi yang akurat, serta memberikan rekomendasi empiris mengenai algoritma mana yang lebih unggul dan efisien untuk diterapkan sebagai sistem prakiraan cuaca lokal yang handal.

2. KAJIAN TEORITIS

Machine Learning (ML) telah bertransformasi menjadi salah satu pilar utama dalam pemodelan prediktif modern, termasuk dalam bidang meteorologi. Konsep dasarnya berfokus pada pengembangan algoritma yang memungkinkan komputer mempelajari pola tersembunyi dari data historis tanpa diprogram secara eksplisit untuk setiap aturan tertentu. Dalam konteks prediksi cuaca, ML menawarkan keunggulan signifikan dibandingkan model statistik tradisional karena kemampuannya menangani hubungan non-linear yang kompleks antarvariabel atmosfer, seperti interaksi antara suhu, kelembapan, dan radiasi surya yang sering kali bersifat dinamis dan fluktuatif (Li et al., 2023). Penelitian (Li et al., 2023) menegaskan bahwa penggunaan algoritma berbasis *ensemble* mampu meningkatkan akurasi prediksi suhu harian secara substansial dibandingkan metode statistik klasik, terutama pada kondisi iklim yang tidak stabil.

Salah satu algoritma yang banyak digunakan adalah *Random Forest* (RF), sebuah metode *ensemble learning* yang diperkenalkan oleh Leo Breiman. RF bekerja dengan membangun sekumpulan pohon keputusan (*decision trees*) selama pelatihan dan mengeluarkan prediksi rata-rata dari pohon-pohon individu tersebut untuk tugas regresi. Kekuatan utama RF terletak pada mekanisme *bootstrap aggregating* (bagging) dan pemilihan fitur acak, yang secara efektif mengurangi varians model dan mencegah masalah *overfitting* yang sering terjadi pada pohon keputusan tunggal (Breiman, 2001). Studi komparatif oleh (Zhang et al., 2022) menemukan bahwa RF memiliki performa yang lebih unggul dan stabil dibandingkan metode *Multivariate Adaptive Regression Splines* (MARS) dalam memprediksi parameter cuaca jangka panjang, menjadikannya kandidat kuat untuk pemodelan iklim lokal.

Di sisi lain, *Support Vector Regression* (SVR) menawarkan pendekatan yang berbeda dengan berbasis pada teori pembelajaran statistik Vapnik. SVR bertujuan untuk menemukan fungsi regresi yang meminimalkan batas kesalahan (ϵ) dari target aktual untuk semua data pelatihan, sekaligus memaksimalkan margin atau kerataan fungsi tersebut. Keunggulan SVR sangat bergantung pada penggunaan teknik *kernel trick*, khususnya *Radial Basis Function* (RBF), yang memungkinkannya memetakan data input dari ruang dimensi rendah ke ruang fitur dimensi tinggi di mana hubungan linear dapat ditemukan (Smaili et al., 2022). (Pal & Foody, 2022) menyoroti bahwa SVR sangat efektif dalam memprediksi parameter lingkungan yang memiliki pola non-linear yang kuat, meskipun kinerjanya sangat sensitif terhadap penyetelan *hyperparameter* seperti C (regularisasi) dan γ (lebar kernel).

Selain pemilihan algoritma, keberhasilan model prediksi cuaca sangat dipengaruhi oleh kualitas fitur input yang digunakan. *Feature Engineering* atau rekayasa fitur merupakan proses mentransformasi data mentah menjadi format yang lebih informatif bagi model. Dalam analisis data deret waktu (*time-series*) meteorologi, variabel suhu hari ini sangat berkorelasi dengan suhu hari-hari sebelumnya. Oleh karena itu, penambahan fitur temporal seperti *lag* (nilai masa lalu) dan *rolling window* (rata-rata bergerak) telah terbukti secara empiris mampu meningkatkan akurasi prediksi dengan menangkap autokorelasi dan tren jangka pendek dalam data (Brownlee, 2020). Berdasarkan landasan teori ini, penelitian ini menghipotesiskan bahwa penerapan *feature engineering* yang tepat pada model SVR yang dioptimalkan akan menghasilkan akurasi prediksi suhu maksimum yang lebih tinggi dibandingkan model *Random Forest* pada dataset cuaca lokal Kota Jambi.

3. METODE PENELITIAN

Penelitian ini menggunakan desain kuantitatif eksperimental untuk mengevaluasi dan membandingkan kinerja model prediksi *Machine Learning*. Populasi penelitian mencakup seluruh data cuaca harian yang tercatat untuk wilayah Kota Jambi, dengan sampel data yang diambil mencakup periode 1 Januari 2020 hingga 31 Desember 2024. Data sekunder ini diperoleh melalui teknik *web scraping* menggunakan API dari platform *Open-Meteo* berdasarkan koordinat geografis spesifik Kota Jambi (1.610°LS, 103.613°BT). Variabel data yang dikumpulkan meliputi suhu maksimum (TMAX) sebagai variabel dependen, serta suhu minimum, curah hujan, kecepatan angin, kelembapan, dan radiasi surya sebagai variabel independen.

Analisis data dilakukan menggunakan perangkat lunak berbasis Python dengan memanfaatkan pustaka *Scikit-learn*. Tahapan pra-pemrosesan data melibatkan teknik imputasi linear untuk menangani *missing values*, serta teknik rekayasa fitur (*Feature Engineering*) untuk membentuk variabel temporal baru. Model *feature engineering* yang diterapkan meliputi pembuatan fitur *Lag* (TMAX_yesterday) dan *Rolling Window* (TMAX_rolling_3d) untuk menangkap pola autokorelasi deret waktu. Data kemudian dinormalisasi menggunakan metode *Z-score standardization* (StandardScaler) sesuai persamaan umum standarisasi statistik, dan dibagi menjadi subset pelatihan (80%) dan pengujian (20%) secara sekuensial.

Model penelitian yang dikembangkan terdiri dari dua algoritma utama: *Random Forest Regressor* dan *Support Vector Regression (SVR)*. SVR diimplementasikan menggunakan kernel *Radial Basis Function (RBF)*, di mana parameter optimal C , ϵ , dan γ ditentukan melalui teknik optimasi *Grid Search*. Evaluasi validitas model tidak dilakukan melalui uji instrumen kuesioner, melainkan melalui validasi silang (*cross-validation*) dan analisis kurva pembelajaran (*learning curve*) untuk mendeteksi *overfitting*. Reliabilitas model diukur berdasarkan konsistensi performa pada data uji menggunakan metrik evaluasi standar regresi, yaitu *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, dan *R-squared (R^2)*, yang merujuk pada formula standar evaluasi regresi (Plevris et al., 2022). Uji signifikansi perbedaan kinerja antar model dilakukan menggunakan uji-t berpasangan (*Paired T-test*) dengan taraf signifikansi $\alpha = 0.05$.

4. HASIL DAN PEMBAHASAN

Hasil analisis statistik deskriptif terhadap variabel target, yaitu suhu maksimum harian (TMAX), mengungkapkan adanya dinamika iklim yang menarik selama periode penelitian. Terlihat adanya tren peningkatan rata-rata suhu harian yang konsisten, di mana rata-rata suhu naik dari 30,03°C pada tahun 2020 menjadi 31,21°C pada tahun 2024. Lebih lanjut, analisis variabilitas data menunjukkan lonjakan suhu ekstrem hingga mencapai 35,2°C pada tahun terakhir, dengan standar deviasi yang semakin melebar. Hal ini mengindikasikan bahwa pola cuaca di Kota Jambi menjadi semakin fluktuatif dan sulit diprediksi, yang menegaskan urgensi penggunaan model non-linear yang kuat.

Hasil Analisis Kinerja Model Prediksi

Berdasarkan eksperimen komputasi yang telah dilakukan, model *Support Vector Regression (SVR)* dan *Random Forest (RF)* dievaluasi menggunakan data uji yang telah melalui proses *feature engineering*. Untuk model SVR, proses optimasi *hyperparameter* menggunakan *Grid Search* menghasilkan konfigurasi optimal berupa kernel RBF dengan parameter $C=10$, $\epsilon=0.2$, dan $\gamma=0.01$.

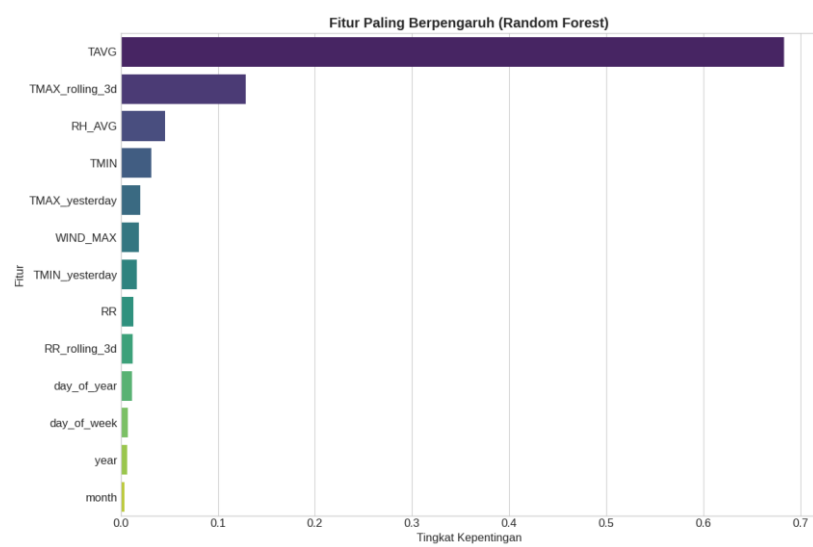
Tabel 1. Perbandingan Kinerja Model Random Forest dan SVR

Metrik Evaluasi	Random Forest (RF)	Support Vector Regression (SVR)
MAE (°C)	0.4339	0.3818
RMSE (°C)	0.5597	0.4964
R2-Score	84.05%	87.46%

Sumber: Hasil Pengolahan Data Peneliti (2025)

Tabel 1 memperlihatkan bahwa model SVR secara konsisten mengungguli *Random Forest* pada seluruh indikator akurasi. Nilai koefisien determinasi (R^2) SVR mencapai 87,46%, yang berarti model ini mampu menjelaskan variasi data suhu aktual sebesar 3,41% lebih baik dibandingkan RF yang hanya mencapai 84,05%. Dari segi presisi, SVR mencatatkan tingkat kesalahan (*error*) yang lebih rendah, dengan RMSE sebesar 0,4964 °C, dibandingkan RF yang memiliki RMSE 0,5597 °C.

Analisis Keterkaitan Fitur dan Konsep Dasar Untuk memahami mekanisme di balik prediksi model, dilakukan analisis tingkat kepentingan fitur (*feature importance*) menggunakan algoritma *Random Forest*, yang hasilnya diilustrasikan pada Gambar 1.



Sumber: Hasil Pengolahan Data Peneliti (2025)

Gambar 1. Grafik pengujung pada suatu website

Gambar 1 menunjukkan bahwa fitur TMAX_yesterday (Lag 1) dan TMAX_rolling_3d merupakan prediktor yang paling dominan. Temuan ini memiliki keterkaitan erat dengan konsep dasar meteorologi mengenai persistensi atmosfer, di mana kondisi cuaca saat ini sangat dipengaruhi oleh inersia termal dari hari-hari sebelumnya. Hal ini membuktikan bahwa keberhasilan model prediksi tidak hanya bergantung pada algoritma, tetapi juga pada kemampuan *feature engineering* dalam merepresentasikan konsep fisik tersebut ke dalam data.

Pembahasan dan Implikasi

Hasil penelitian ini menunjukkan kesesuaian dengan teori dasar SVR yang dikemukakan oleh Vapnik, di mana penggunaan kernel RBF efektif untuk memetakan data non-linear ke dimensi tinggi (Smaili et al., 2022). Kemampuan SVR mengungguli RF dalam studi ini sejalan dengan temuan (Karnisih et al., 2025) di wilayah Malang, yang juga melaporkan superioritas SVR dalam menangani fluktuasi suhu mikro. Namun, hasil ini bertentangan dengan studi

(Indrasari & Suhendar, n.d.) di Bogor yang mengunggulkan *Random Forest*. Perbedaan ini mengindikasikan bahwa efektivitas model *machine learning* bersifat spesifik terhadap lokasi (*site-specific*) di Kota Jambi yang memiliki tren kenaikan suhu ekstrem dan variabilitas tinggi, pendekatan *hyperplane* kontinu SVR terbukti lebih adaptif dibandingkan pendekatan partisi ruang diskrit milik RF.

Secara teoritis, penelitian ini memberikan kontribusi bukti empiris mengenai pentingnya fitur temporal (*lag* dan *rolling*) dalam meningkatkan akurasi model regresi untuk data iklim tropis. Secara terapan, implikasi dari temuan ini adalah rekomendasi penggunaan model SVR yang dioptimalkan sebagai mesin inferensi utama dalam pengembangan sistem peringatan dini cuaca panas ekstrem di Kota Jambi. Sistem ini dapat membantu sektor pertanian dalam menjadwalkan irigasi dan sektor kesehatan dalam mengantisipasi penyakit terkait cuaca panas.

5. KESIMPULAN DAN SARAN

Penelitian ini menyimpulkan bahwa algoritma *Support Vector Regression* (SVR) dengan optimasi kernel RBF merupakan metode yang lebih efektif dibandingkan *Random Forest* untuk memprediksi suhu maksimum harian di Kota Jambi. Berdasarkan hasil analisis kinerja, SVR mampu menghasilkan akurasi prediksi yang lebih tinggi dengan nilai R^2 mencapai 87,46% dan tingkat kesalahan (RMSE) sebesar 0,4964 °C, mengungguli *Random Forest* yang mencapai R^2 84,05%. Uji hipotesis statistik mengonfirmasi signifikansi keunggulan SVR, yang menunjukkan kemampuannya dalam menangani pola non-linear pada data iklim tropis. Selain itu, temuan ini menegaskan bahwa penerapan rekayasa fitur temporal, khususnya fitur *lag* dan *rolling window*, memegang peranan krusial dalam keberhasilan model dalam menangkap sifat persistensi atmosfer lokal.

Berdasarkan kesimpulan tersebut, direkomendasikan untuk menerapkan model SVR yang dioptimalkan sebagai landasan pengembangan sistem prakiraan cuaca lokal guna mendukung mitigasi dampak cuaca panas. Namun, penelitian ini memiliki keterbatasan pada cakupan data historis yang hanya lima tahun dan variabel meteorologis yang terbatas. Oleh karena itu, penelitian masa depan sangat disarankan untuk memperluas rentang periode data guna menangkap siklus iklim jangka panjang, serta mengeksplorasi penambahan variabel eksternal seperti tekanan udara dan indeks iklim global. Studi lanjutan yang membandingkan SVR dengan algoritma *Deep Learning* seperti LSTM (*Long Short-Term Memory*) juga direkomendasikan untuk melihat potensi peningkatan akurasi pada data deret waktu yang kompleks.

UCAPAN TERIMA KASIH

Penulis menyampaikan penghargaan dan ucapan terima kasih yang tulus kepada Universitas Dinamika Bangsa Jambi atas dukungan fasilitas laboratorium dan akses terhadap sumber daya yang memungkinkan penelitian ini dapat terlaksana dengan baik. Selain itu, penulis mengapresiasi platform *Open-Meteo* atas penyediaan akses data meteorologi historis yang menjadi komponen fundamental dalam studi ini.

DAFTAR REFERENSI

- Béland, H. K. & Sun, Y. (2021). Machine Learning in Atmospheric Science: A Review. *Atmosphere*, 12(8). <https://doi.org/10.3390/atmos12080988>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.
- Brownlee, J. (2020). *Machine Learning Mastery With Python*. Machine Learning Mastery.
- Gupta, A., Maji, S. & Debnath, R. (2022). Machine Learning Approaches for Weather Prediction: An Overview. *Environmental Modelling & Software*, 152. <https://doi.org/10.1016/j.envsoft.2022.105360>
- Indrasari, W. & Suhendar, H. (n.d.). ANALISIS MODEL PREDIKSI CUACA MENGGUNAKAN SUPPORT VECTOR MACHINE, GRADIENT BOOSTING, RANDOM FOREST, DAN DECISION TREE. *Prosiding Seminar Nasional Fisika (E-Journal)*, XII. <https://doi.org/10.21009/03.1201.FA18>
- Kumar, V. & others. (2021). A Review on Numerical Weather Prediction Models and Their Applications. *Earth-Science Reviews*, 221. <https://doi.org/10.1016/j.earscirev.2021.103774>
- Li, Y., Li, T., Lv, W., Liang, Z. & Wang, J. (2023). Prediction of Daily Temperature Based on the Robust Machine Learning Algorithms. *Sustainability (Switzerland)*, 15(12). <https://doi.org/10.3390/su15129289>
- Pal, M. & Foody, D. (2022). Applications of Support Vector Machines in the Prediction of Environmental Parameters. *Environmental Modelling & Software*, 150. <https://doi.org/10.1016/j.envsoft.2022.105296>
- Plevris, V., Solorzano, G., Bakas, N. P. & Ben Seghier, M. E. A. (2022). INVESTIGATION OF PERFORMANCE METRICS IN REGRESSION ANALYSIS AND MACHINE LEARNING-BASED PREDICTION MODELS. *World Congress in Computational Mechanics and ECCOMAS Congress*. <https://doi.org/10.23967/eccomas.2022.155>
- Smaili, A., Seridi, H. & Sandli, F. E. (2022). Support Vector Regression Models for Nonlinear System Modeling: A Comprehensive Review. *Neural Computing and Applications*, 34, 10341–10363. <https://doi.org/10.1007/s00521-021-06735-w>
- Smola, A. J. & Schölkopf, B. (2004). A Tutorial on Support Vector Regression. *Statistics and Computing*, 14(3), 199–222.
- Xu, W. & others. (2022). Trends and Perspectives in Artificial Intelligence and Machine Learning for Environmental Data Analysis. *Environmental Science & Technology*, 56(12), 8197–8212. <https://doi.org/10.1021/acs.est.1c08499>

Zhang, G., Bateni, S. M., Jun, C., Khoshkam, H., Band, S. S. & Mosavi, A. (2022). Feasibility of Random Forest and Multivariate Adaptive Regression Splines for Predicting Long-Term Mean Monthly Dew Point Temperature. *Frontiers in Environmental Science*, 10. <https://doi.org/10.3389/fenvs.2022.826165>