JISTech (Journal of Islamic Science and Technology)

JISTech, 10(1), 46-55, January – June 2025

ISSN: 2528-5718

http://jurnal.uinsu.ac.id/index.php/jistech

Prediksi Tinggi Muka Air Di Lahan Gambut Provinsi Jambi Berdasarkan Curah Hujan, Suhu Dan Kelembapan Udara Menggunakan Metode Random Forest Regression

Randi Irawan[1*], Gusmi Kholijah[2], Cut Multahadah[3]

[1,2,3] *Universitas Jambi, Jambi, Indonesia*

ABSTRAK

Lahan gambut merupakan ekosistem penting yang rentan terhadap kerusakan, terutama akibat kebakaran yang sering dipicu oleh penurunan tinggi muka air (TMA). Provinsi Jambi, sebagai salah satu wilayah dengan luasan gambut yang signifikan, mengalami kebakaran berulang dalam beberapa tahun terakhir. Penelitian ini bertujuan untuk membangun model prediksi TMA berbasis algoritma Random Forest Regression guna mendukung mitigasi kebakaran dan pengelolaan lahan gambut berkelanjutan. Model dilatih menggunakan data iklim (curah hujan, suhu, dan kelembaban udara) dari tahun 2019 hingga 2024 di wilayah gambut Provinsi Jambi, dengan 100 pohon keputusan dan pendekatan bootstrap sampling, menggunakan Python dan pustaka Scikit-Learn. Hasil evaluasi menunjukkan performa model yang cukup baik, dengan nilai RMSE sebesar 0,1173 dan R^2 sebesar 0,7182, yang berarti sekitar 72% variasi TMA dapat dijelaskan oleh model. Analisis feature importance mengidentifikasi curah hujan sebagai faktor paling berpengaruh yaitu sebesar 51.05%, disusul kelembaban sebesar 32.9% dan suhu sebesar 11.9%. Temuan ini menegaskan peran utama curah hujan dalam dinamika TMA, serta potensi penggunaan model sebagai alat bantu dalam sistem peringatan dini dan perencanaan pengelolaan lahan gambut yang lebih efektif.

ABSTRACT

Peatlands are important ecosystems that are vulnerable to damage, especially due to fires that are often triggered by low water levels (TMA). Jambi Province, as one of the areas with significant peat areas, has experienced repeated fires in recent years. This study aims to build a TMA prediction model based on the Random Forest Regression algorithm to support fire mitigation and sustainable peatland management. The model was trained using daily climate data (rainfall, temperature, and air humidity) from 2019 to 2024 in the peat area of Jambi Province, with 100 Decision Trees and a bootstrap sampling approach, using Python and the Scikit-Learn library. The results of the evaluation showed a fairly good model performance, with an RMSE value of 0.1173 and an R^2 of 0.7182, which means that about 72% of the TMA variation can be explained by the model. The feature importance analysis identified rainfall as the most influential factor at 51.05%, followed by humidity at 32.9% and temperature at 11.9%. These findings confirm the key role of rainfall in the dynamics of TMA, as well as the potential use of the model as a tool in early warning systems and more effective peatland management planning.

Kata Kunci: Curah hujan, Iklim, Kelembaban udara, Lahan gambut, Random Forest Regression, Suhu, Tinggi muka air.
Email: * irawanrandi109@gmail.com

DOI: http://dx.doi.org/10.30829/jistech.v10i1.24569
Diterima 9 Maret 2025; Direvisi 15 Juni 2025; Disetujui 25 Juni 2025

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Pendahuluan

Gambut merupakan jenis tanah yang terbentuk dari akumulasi bahan organik seperti sisa-sisa tumbuhan yang membusuk secara perlahan dalam kondisi basah dan minim oksigen selama ribuan tahun. Lapisan tanah ini kaya akan karbon dan berfungsi sebagai penyimpan cadangan karbon terbesar di daratan. Gambut juga berperan penting dalam menjaga keseimbangan alam melalui kemampuannya menyerap dan menyimpan air, mencegah banjir dan kekeringan, serta menjadi habitat bagi berbagai jenis flora dan fauna endemik. Sifat gambut yang menyimpan air membuatnya sangat bergantung pada kondisi hidrologis. Ketika lahan gambut kering tanah ini menjadi sangat mudah terbakar. Kebakaran lahan di Indonesia yang terjadi dalam kurun waktu tahun 2015-2019 mencapai 4,4 juta ha dengan 50% diantaranya merupakan lahan gambut, bahkan sekitar 18% dari lahan yang terbakar mengalami kebakaran berulang [1]. Laporan World Bank tahun 2016 menyampaikan bahwa kebakaran hutan dan lahan tahun 2015 menjadi salah satu kebakaran terbesar di Indonesia pasca kebakaran tahun 1997/1998, yang

mengakibatkan kerugian global hingga Rp 221 triliun atau setara 16,1 miliar dolar AS. Lalu Badan Nasional Penanggulangan Bencana (BNPB) juga mencatat setengah juta orang masuk rumah sakit serta puluhan orang meninggal karena kebakaran lahan [1].

Wahana Lingkungan Hidup Indonesia (WALHI) mengemukakan bahwa tahun 2015-2019 menjadi catatan kelam bagi Provinsi Jambi dalam rentang tahun tersebut terjadi kebakaran seluas 238.401,3 ha [2]. Luasan tersebut 67% (158.971 ha) areal terbakar berada pada areal gambut dan sisanya 33% (79.430,3 ha) pada areal mineral. Data Dinas Kehutanan Provinsi Jambi menyampaikan 80% dari 900 ribu hektare total luas lahan gambut di Jambi rusak akibat kebakaran pada tahun 2015. Kondisi ini semakin parah karena kebakaran terus berulang di tahun selanjutnya. Besarnya angka kebakaran lahan gambut di Provinsi Jambi terjadi karena Provinsi Jambi memiliki luas area lahan gambut mencapai 736.227,20 ha atau sekitar 14% dari luas Provinsi Jambi [3]. Badan Riset dan Inovasi Nasional (BRIN) tahun 2024 menyampaikan bahwa salah satu faktor utama yang berkontribusi terhadap kebakaran lahan gambut adalah rendahnya tinggi muka air. Pemerintah telah menerbitkan berbagai regulasi dalam upaya untuk melindungi ekosistem gambut, seperti Peraturan Pemerintah Nomor 57 Tahun 2016 tentang Perlindungan dan Pengelolaan Ekosistem Gambut. Ekosistem gambut dengan fungsi budidaya dinyatakan rusak apabila tinggi muka air tanah lebih dari - 0,4 meter di bawah permukaan gambut.

Tinggi muka air mengalami kenaikan dan penurunan salah satunya dipengaruhi oleh iklim. Perubahan iklim terhadap potensi kebakaran menunjukkan bahwa perubahan iklim berpengaruh terhadap penurunan tinggi muka air yang mengakibatkan bertambahnya jumlah titik panas di lahan gambut [4]. Perubahan iklim berpengaruh terhadap penurunan kelembaban tanah gambut yang dipicu oleh rendahnya tinggi muka air. Unsur iklim yang menpengaruhi yaitu curah hujan, suhu dan kelembaban udara. Curah hujan yang tinggi menambah pasokan air ke permukaan tanah dan kanal. Lalu suhu udara yang tinggi dapat mempercepat penguapan, yang berpotensi menurunkan tinggi muka air [3]. Kemudian kelembaban udara yang tinggi justru dapat memperlambat proses pengeringan lahan sehingga mempertahankan kadar air lebih lama [4]. Kombinasi unsur iklim yang mempengaruhi tinggi muka air membentuk pola hubungan yang kompleks, dengan efek masing-masing variabel saling bergantung dan tidak selalu mengikuti pola tetap.

Analisis tinggi muka air memerlukan metode yang mampu menangani kompleksitas hubungan antar variabel yang tidak linier dan saling mempengaruhi. Metode konvensional sering kali kurang efektif dalam mengakomodasi faktor-faktor tersebut, sehingga diperlukan pendekatan yang lebih adaptif dan akurat. *Random forest regression* menjadi salah satu metode yang dapat digunakan karena kemampuannya dalam menangani variabel yang kompleks [5]. *Random forest regression* untuk mengestimasi kelembapan tanah gambut dengan memanfaatkan data observasi kelembapan tanah sebagai target dan data satelit seperti curah hujan (CHIRPS), serta *Normalized Difference Vegetation Index* (NDVI), *Soil Adjusted Vegetation Index* (SAVI), *Normalized Difference Moisture Index* (NDMI*), Normalized Difference Soil Index* (NDSI), dan *Normalized Difference Water Index* (NDWI) yang diekstrak dari satelit Landsat 8 sebagai data prediktor. Pada penelitian ini model *Random forest regression* yang dikembangkan menunjukkan kinerja yang baik dalam memprediksi kelembapan tanah. Penelitian lainnya oleh Firmansyah et al., (2024) mengenai algoritma yang *Random forest regression* digunakan untuk prediksi kebakaran lahan gambut di kabupaten Okan Komering Ilir Provinsi Sumatera Selatan, faktor-faktor yang digunakan untuk memprediksi kebakaran hutan meliputi data titik panas, yang mengidentifikasi lokasi-lokasi berpotensi terjadi kebakaran, serta data iklim yang mencakup curah hujan, suhu, kelembapan, dan sinar matahari. Pada penelitian ini algoritma *Random forest regression* efektif dalam mengklasifikasikan data kebakaran, dengan hasil evaluasi yang menunjukkan kinerja yang baik.

Metode *Random forest regression* bekerja dengan membangun banyak pohon keputusan *(Decision Trees)* yang masing-masing membagi data berdasarkan aturan tertentu [7]. Pohon keputusan dalam penelitian ini akan mengevaluasi faktor-faktor seperti curah hujan, suhu, kelembaban udara berpengaruh terhadap tinggi muka air. Banyaknya pohon keputusan yang dibangun oleh metode *Random forest* dapat menangkap pola hubungan antar variabel secara lebih akurat [8]. Metode *Random forest* juga mampu mengidentifikasi faktor yang paling berpengaruh terhadap tinggi muka air [9]. Berdasarkan latar belakang, penelitian ini mengembangkan model prediksi tinggi muka air berdasarkan data curah hujan, kelembaban udara dan suhu menggunakan metode *random forest.*

Metodologi Penelitian

a. Jenis dan Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari dua sumber yaitu:

1) Data Tinggi Muka Air (TMA), diperoleh dari Badan Restorasi Gambut dan Mangrove (BRGM). Data ini mencakup pengukuran tinggi muka air di lahan gambut Provinsi Jambi dalam rentang waktu 2019 s/d 2024 dan pada berbagai titik pemantauan tiga kabupaten yaitu Kabupaten Muaro Jambi, Tanjung Jabung Timur dan Tanjung Jabung Barat
2) Data Iklim, (curah hujan, suhu, dan kelembapan udara) diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Data ini digunakan sebagai variabel prediktor dalam model prediksi tinggi muka air guna mengidentifikasi pola hubungan antara kondisi iklim dan perubahan tinggi muka air di lahan gambut.

b. Variabel Penelitian

Penelitian ini menggunakan beberapa variabel yang berperan dalam prediksi tinggi muka air di lahan

gambut. Variabel penelitian dapat dilihat pada tabel berikut:

Tabel 1. Variabel Penelitian

Variabel	Notasi	Definisi Variabel	Skala Data	Satuan
Tinggi Muka Air (TMA)	y	Kedalaman muka air yang di ukur dari permukaan tanah gambut	Rasio	Meter (m)
Suhu	x_1	Suhu rata-rata bulanan di wilayah penelitian	Rasio	Derajat Celsius (°C)
Kelembapan udara	x_2	Persentase kelembapan udara bulanan	Rasio	Persentase (%)
Curah hujan	x_3	Rata-rata jumlah curah hujan yang diterima setiap bulanya	Rasio	Milimeter (mm)

c. Teknik Analisis Data

Analisis data dalam penelitian ini dilakukan melalui beberapa tahap utama, yaitu:

1) Pra-pemrosesan data
   a) Penggabungan data, data Iklim (curah hujan, suhu, kelembapan) dan Tinggi Muka air digabungkan berdasarkan rentang waktu yang sama agar memiliki keterkaitan antar variabel.
   b) Pembersihan data, mengisi data yang hilang (*missing values*) atau melakukan imputasi menggunakan rata-rata.
   c) Pembagian data *training* dan *data testing,* Membagi data penelitian menjadi dua bagian yaitu data *training* dan data *testing* dengan perbandingan 80:20.
2) Penerapan metode *Random forest regression*

   Metode *Random forest regressionRegression* digunakan untuk membangun model prediksi tinggi muka air berdasarkan data iklim.

   Tahapan Penerapan *Random forest regression:*
   1. *Bootstrap Sampling*

      Data pelatihan diambil secara acak dengan rasio tertentu menggunakan teknik bootstrap. Setiap subset digunakan untuk melatih satu pohon keputusan regresi. Jumlah bootstrap sesuai dengan jumlah pohon yang ditentukan.
   2. Pembuatan Pohon Keputusan

      Fitur curah hujan (mm), suhu (°C), dan kelembaban udara (%) digunakan untuk menentukan split terbaik pada setiap node. Kedalaman pohon dibatasi hingga level 3 (max_depth=3) guna menghindari overfitting dan menjaga interpretabilitas. Pembatasan ini sesuai rekomendasi untuk dataset dengan kurang dari 100 sampel (Rian et al., 2024).

      Proses dimulai dengan menghitung rata-rata target (TMA) pada setiap node dan nilai Mean Squared Error (MSE) sebagai ukuran ketidakhomogenan. Untuk setiap fitur, ditentukan beberapa threshold sebagai kandidat split. Masing-masing threshold diuji dengan membagi data menjadi dua subset dan menghitung MSE total setelah split (rata-rata tertimbang dari kedua subset). Gain dihitung dari selisih MSE sebelum dan sesudah split. Split terbaik adalah yang memberikan gain terbesar. Proses ini berlanjut secara rekursif hingga tercapai kedalaman maksimum atau kriteria berhenti lainnya.
   3. Model Pohon Keputusan

      Setiap pohon dibentuk berdasarkan struktur hasil proses split. Prediksi akhir dari pohon diperoleh dari rata-rata target pada masing-masing daun. Secara matematis, pohon dapat dinyatakan sebagai fungsi potongan (piecewise function) berdasarkan aturan logika fitur prediktor.
   4. *Agregatting*

      Prediksi akhir Random Forest diperoleh dengan menghitung rata-rata dari seluruh prediksi pohon.
   5. Prediksi Data Uji, Model yang telah dibangun digunakan untuk memprediksi data uji. Setiap pohon memberikan satu prediksi per sampel, lalu rata-ratanya dihitung sebagai prediksi akhir.
3) Evaluasi model

   Evaluasi model dilakukan dengan menggunakan metrik *Root Mean Squared Error (RMSE)* dan *r-squared (R^2)* untuk mengukur performa model dalam memprediksi tinggi muka air dengan akurasi yang memadai. Data uji memberikan gambaran objektif tentang bagaimana model akan bekerja dengan data yang baru dan belum pernah digunakan sebelumnya.
4) *Feature importance*

   Secara umum *Feature importance* untuk sebuah fitur dihitung dengan cara:
   1. Menjumlahkan penurunan *impurity* yang terjadi setiap kali fitur tersebut digunakan untuk melakukan *split* dalam suatu tree.
   2. Mengalikan penurunan *impurity* tersebut dengan proporsi sampel yang melewati *node* tempat *split* dilakukan.
   3. Menjumlahkan kontribusi dari semua *node* dalam satu tree.

Hasil dan Pembahasan

Hasil dari penerapan model *Random forest regression* yang telah dibangun dalam penelitian ini dimulai dengan menggambarkan hasil dari tahapan-tahapan yang telah dilakukan, mulai dari pra-pemrosesan data, proses *bootstrap*, pembangunan pohon keputusan, hingga evaluasi model yang diukur menggunakan metrik-metrik evaluasi yaitu *Root Mean Squared Error* (R*MSE*) dan *koefisien determinasi* (R^2).

a. Pre-processing data

Tahap pertama adalah proses *preprocessing* data. *Preprocessing* dilakukan untuk memastikan bahwa data yang digunakan dalam pemodelan memiliki kualitas yang baik dan siap untuk dianalisis menggunakan algoritma machine learning. Proses ini mencakup beberapa langkah utama sebagai berikut:

1) Pemeriksaan missing values

Sebelum melatih model *Random forest regression*, satu langkah yang tak bisa diabaikan adalah menangani *missing values* dalam dataset. Algoritma *Random forest*, yang mengandalkan pembuatan banyak pohon keputusan, secara fundamental tidak mampu memproses data yang tidak lengkap. Keberadaan nilai yang hilang akan mengganggu kemampuan model untuk mengidentifikasi pola dan hubungan dalam data, bahkan bisa mencegah proses pelatihan model itu sendiri. Maka memastikan data bersih dan lengkap dari *missing values* adalah prasyarat mutlak untuk membuat model *Random forest regression*. Berikut merupakan hasil informasi dari *missing values* dapat dilihat pada tabel 3.

Tabel 2. *Missing Values*

Variabel	Analisis *Missing Values*
Suhu (x_1)	0
Kelembaban Udara (x_2)	0
Curah Hujan (x_3)	0
Tinggi Muka Air (TMA) (y)	0

Setelah melakukan analisis terhadap dataset yang tersedia, dapat disimpulkan bahwa tidak ditemukan adanya nilai yang hilang (*missing values*) pada semua variabel yang dianalisis, yaitu curah hujan, suhu, kelembaban udara, dan TMA, yang semuanya memiliki jumlah missing values sebesar 0. Informasi ini menunjukkan bahwa dataset sudah lengkap tanpa perlu dilakukan imputasi atau pengisian data terkait *missing values* artinya analisis selanjutnya dapat dilanjutkan.

2) Pembagian data *training* dan data *testing*

Pada penelitian ini pembagian *dataset* dilakukan dengan proporsi 80% untuk data latih (*training*) dan 20% untuk data uji (*testing*). Pembagian data ini bertujuan untuk memastikan model memperoleh data yang memadai guna mempelajari pola secara efektif dari data latih, sekaligus memungkinkan evaluasi kinerja yang objektif menggunakan data uji yang belum pernah dikenali sebelumnya. Berdasarkan penelitian oleh Mohammed dan Alsunosi (2022), rasio pembagian 80:20 dapat memberikan performa optimal untuk model regresi, khususnya pada *dataset* berukuran kecil hingga sedang, dalam penelitiannya diperoleh bahwa peningkatan proporsi data latih secara signifikan meningkatkan kinerja model *Random forest regressionRegression*, dengan rasio 80:20 memberikan keseimbangan ideal antara kapabilitas pembelajaran model dan validitas evaluasi.

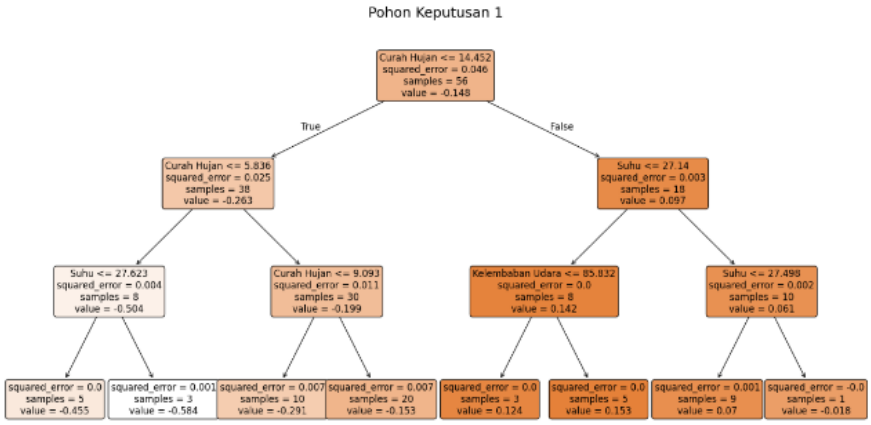
b. *Random Forest Regression*

1) *Bootstrap sampling*

Bootstrap sampling dilakukan untuk mengambil sekumpulan sampel dari dataset. Bootstrap ini melibatkan pengambilan sampel secara acak dengan pengembalian (*sampling with replacement*) dari dataset. Pada penelitian ini dibuat 100 *subset* hasil *bootstrap sampling*, yang berarti jumlah pohon keputusan yang akan dibangun juga berjumlah 100.

2) *Pembuatan Pohon Keputusan (Decision Tree)*

Setiap *subset* hasil *bootstrap* sebanyak 100 subset digunakan untuk membangun satu pohon keputusan (*Decision Tree*), sehingga terbentuk 100 pohon sesuai dengan jumlah *subset bootstrap*. Setiap pohon terdiri dari *node* akar, *node* cabang, dan *node* daun, yang dibentuk melalui proses pemilihan *split* terbaik berdasarkan nilai *Information gain terbesar.* Pada penelitian ini penulis membatasi maksimal kedalaman pohon (maxdepth) sebesar 3 yang artinya pohon keputusan memiliki 3 tingkat node yang bercabang sebelum node daun. Pembatasan ini dilakukan agar setiap pohon memiliki struktur yang konsisten. Representasi visual dari pohon keputusan yang terbentuk dengan kedalaman maksimum 3 level (*maxdepth*=3), dimana setiap node cabang menampilkan informasi *threshold*, nilai *MSE*, dan jumlah sampel yang terdapat pada node tersebut, sedangkan node daun menampilkan nilai *MSE*, jumlah sampel, dan nilai rata-rata sebagai hasil prediksi akhir. Visualisasi salah satu pohon keputusan dalam *Random Forest Regression* berikut ini:



Gambar 1. Pohon keputusan (Decision Tree)

Berdasarkan visualisasi pohon keputusan diatas, proses pemisahan data diawali dari node akar dengan menggunakan variabel Curah Hujan sebagai kriteria pemisahan. Nilai *threshold* yang digunakan pada node ini adalah Curah Hujan ≤ 14.452 mm, yang menghasilkan *MSE* sebesar 0.046 dengan jumlah sampel sebanyak 56 serta rata-rata nilai target sebesar -0.148 m. Pemilihan Curah Hujan sebagai variabel pemisah utama menunjukkan bahwa variabel ini memiliki kontribusi terbesar dalam mengurangi varians prediksi pada tahap awal pembentukan pohon ke-1.

Pada cabang kiri dari node akar (kondisi Curah Hujan ≤ 14.452 mm), dilakukan pemisahan lanjutan dengan menggunakan Curah Hujan ≤ 5.836 mm sebagai *threshold*. Split ini menghasilkan *MSE* sebesar 0.025 dengan 38 sampel dan rata-rata nilai target sebesar -0.263 m. Cabang ini kemudian dibagi ke dalam dua sub-cabang, cabang kiri menggunakan Suhu ≤ 27.623 °C sebagai *threshold* yang menghasilkan node daun kiri dengan 5 sampel, rata-rata nilai target sebesar -0.455 m, serta *MSE* sebesar 0.0001 dan node daun kanan dengan 3 sampel, rata-rata nilai target sebesar -0.584 m, serta *MSE* sebesar 0.001. Selanjutnya cabang kanan menggunakan Curah Hujan ≤ 9.093 mm sebagai *threshold*, menghasilkan node daun kiri dengan 10 sampel dan rata-rata nilai target sebesar -0.291 m serta *MSE* sebesar 0.007, dan node daun kanan dengan 20 sampel, rata-rata nilai target sebesar -0.153 m, serta *MSE* sebesar 0.007.

Pada cabang kanan dari node akar (kondisi Curah Hujan > 14.452 mm), pemisahan dilakukan berdasarkan variabel Suhu ≤ 27.14 °C sebagai *threshold*, dengan *MSE* sebesar 0.003, 18 sampel, dan rata-rata nilai target sebesar 0.097 m. Cabang ini terbagi menjadi dua bagian. Cabang kiri menggunakan Kelembaban Udara ≤ 85.832 % sebagai *threshold*, yang menghasilkan node daun kiri dengan 5 sampel, rata-rata nilai target sebesar 0.153 m, serta *MSE* sebesar 0.0001, dan node daun kanan dengan 3 sampel, rata-rata nilai target sebesar 0.124 m, serta *MSE* sebesar 0.0001. Kemudian cabang kanan menggunakan Suhu ≤ 27.498 °C sebagai *threshold*, menghasilkan node daun kiri dengan 9 sampel, rata-rata nilai target sebesar 0.070 m, serta *MSE* sebesar 0.001, dan node daun kanan dengan 1 sampel, rata-rata nilai target sebesar -0.018 m, serta *MSE* sebesar 0.0001.

Secara keseluruhan struktur pohon keputusan ini menunjukkan bahwa pemisahan berdasarkan *threshold* secara bertahap mampu meningkatkan homogenitas dalam kelompok data, sebagaimana tercermin dari penurunan nilai *MSE* dari node akar hingga node daun. Pola pemisahan yang terbentuk mengindikasikan bahwa nilai target cenderung negatif pada kondisi Curah Hujan rendah dan positif pada kondisi Curah Hujan tinggi, dengan variabel Suhu dan Kelembaban Udara sebagai faktor pendukung dalam pemisahan di level yang lebih dalam.

3) *Model Decision Tree*

Secara matematis, bentuk model dari Decision Tree berdasarkan pohon keputusan ke-1 dapat direpresentasikan dalam bentuk fungsi bersyarat yang menggambarkan proses pengambilan keputusan pada setiap node. Model ini mengikuti struktur hierarkis yang telah dijelaskan sebelumnya dimana setiap percabangan didasarkan pada *threshold* tertentu dari fitur-fitur yang dipilih.

Model-model dari pohon-pohon keputusan lainnya dalam *Random forest regression* yang masing-masing dibangun berdasarkan *subset bootstrap* yang berbeda, representasi matematis serupa dapat diterapkan dengan struktur dan parameter yang disesuaikan dengan karakteristik masing-masing pohon. Berikut ini adalah bentuk model dari pohon keputusan ke-1 pada *subset bootstrap* ke-1.

$$\hat{y}^{(1)} = \begin{cases} -0.455 & jika\ x_3 \leq 14.452\ \wedge\ x_3 \leq 5.836\ \wedge\ x_1 \leq 27.623 \\ -0.584 & jika\ x_3 \leq 14.452\ \wedge\ x_3 \leq 5.836\ \wedge\ x_1 > 27.623 \\ -0.291 & jika\ x_3 \leq 14.452\ \wedge\ x_3 > 5.836\ \wedge\ x_3 \leq 9.093 \\ -0.153 & jika\ x_3 \leq 14.452\ \wedge\ x_3 > 5.836\ \wedge\ x_3 > 9.093 \\ 0.124 & jika\ x_3 > 14.452\ \wedge\ x_1 \leq 27.140\ \wedge\ x_2 \leq 85.832 \\ 0.153 & jika\ x_3 > 14.452\ \wedge\ x_1 \leq 27.140\ \wedge\ x_2 > 85.832 \\ 0.070 & jika\ x_3 > 14.452\ \wedge\ x_1 > 27.140\ \wedge\ x_1 \leq 27.498 \\ -0.018 & jika\ x_3 > 14.452\ \wedge\ x_1 > 27.140\ \wedge\ x_1 > 27.498 \end{cases}$$

Berdasarkan model dari pohon keputusan ke-1 diperoleh bahwa prediksi tinggi muka air $y^{(1)}$ ditentukan oleh nilai Curah Hujan (x_3), Suhu (x_1), dan Kelembaban Udara (x_2). Ketika curah hujan memenuhi kondisi $x_3 \leq 5{,}836$ dan suhu $x_1 \leq 27{,}623$ °C, maka model memprediksi tinggi muka air sebesar $-0{,}455$ m. Jika suhu melebihi ambang tersebut, yaitu $x_1 > 27{,}623$ °C, maka nilai prediksi menjadi $-0{,}584$. Selanjutnya jika curah hujan berada dalam rentang $5.836\ mm < x_3 \leq 14.452\ mm$, maka model mempertimbangkan *threshold* lanjutan. Jika $x_3 \leq 9{,}093\ mm$, maka prediksi TMA adalah $-0{,}291$ m, sedangkan jika $x_3 > 9{,}093\ mm$, maka prediksi TMA adalah $-0{,}153\ m$.

Pada kasus ketika curah hujan melebihi ambang utama, yaitu $x_3 > 14.452\ mm$, maka pemisahan dilakukan berdasarkan nilai suhu. Jika suhu $x_1 \leq 27.140$ °C, maka model melanjutkan pemisahan berdasarkan kelembaban udara. Ketika $x_2 \leq 85.832$, maka nilai prediksi adalah 0,124 m, jika kelembaban udara $x_2 > 85.832$ %, maka nilai prediksi meningkat menjadi 0,153 m. Kemudian apabila suhu $x_1 > 27{,}140$ °C, maka pohon mengevaluasi kembali nilai suhu, jika $x_1 \leq 27.492$ °C, maka nilai prediksi adalah 0,070 m, sedangkan jika $x_1 > 27{,}498$ °C, maka model memberikan prediksi sebesar $-0{,}018$ m.

Model pohon keputusan ini secara keseluruhan menghasilkan delapan kondisi akhir yang ditentukan oleh kombinasi nilai *threshold* dari variabel x_3, x_1, dan x_2. Setiap jalur keputusan mewakili kombinasi kondisi *True* atau *False* terhadap masing-masing *threshold* yang berujung pada satu node daun dengan nilai prediksi tertentu. Secara umum nilai tinggi muka air cenderung negatif pada kondisi curah hujan rendah ($x_3 \leq 14.452\ mm$), dan menjadi positif saat curah hujan tinggi ($x_3 > 14.452\ mm$) dikombinasikan dengan suhu dan kelembaban udara. Struktur pohon ini menunjukkan kemampuan dalam memetakan hubungan non-linear antar variabel input terhadap variabel target, dengan akurasi yang ditunjukkan melalui nilai *Mean Squared Error* (*MSE*) yang semakin kecil pada tiap level pemisahan.

4) *Prediksi akhir (agregatting) menggunakan Data Testing*

Setelah model dilatih dengan data *training*, evaluasi dilanjutkan dengan menggunakan data *testing* yang belum pernah dilihat oleh model sebelumnya. Setiap model pohon keputusan dalam ensemble *Random forest regression* menghasilkan prediksi untuk sampel data *testing*, kemudian prediksi tersebut diagregasi dengan menghitung rata-rata untuk mendapatkan prediksi akhir. Proses ini sama seperti pada proses *agregatting* menggunakan data training pada subbab sebelumnya dengan memeriksa model yang memenuhi kriteria sampel pada data *testing*. Pada sampel ke-1 data testing yaitu suhu ($x_1 = 27.121$), kelembaban udara ($x_2 = 86.414$), dan curah hujan ($x_3 = 12.523$). Ketiga variabel prediktor tersebut digunakan oleh model untuk menentukan jalur keputusan dalam pohon keputusan.

Pertama, model mengevaluasi apakah $x_3 \leq 14{,}452$ karena nilai x_3 sebesar 12,523 memenuhi kondisi tersebut (*True*), maka jalur dilanjutkan ke cabang kiri, selanjutnya model kembali mengevaluasi apakah $x_3 \leq 5{,}836$ karena 12,523 lebih besar dari 5,836 (*False*), maka jalur bergeser ke cabang kanan. Pada cabang ini model memeriksa *threshold* lanjutan, yaitu apakah $x_3 \leq 9{,}093$ karena nilai x_3 sebesar 12,523 lebih besar dari 9,093 (*False*), maka jalur kembali bergerak ke kanan. Berdasarkan struktur pohon, kondisi ini membawa model ke node daun yang memuat nilai prediksi sebesar $-0{,}153$.

Sehingga dari model pohon keputusan ke-1, prediksi tinggi muka air untuk sampel ke-1 data testing adalah $-0{,}153$. Karena *Random forest regression* terdiri dari 100 pohon, prediksi akhir akan diperoleh dengan merata-ratakan hasil dari semua pohon yang memberikan prediksi untuk sampel tersebut. Prediksi $-0{,}153$ dari model pohon pertama menjadi salah satu kontribusi terhadap hasil akhir tersebut. Proses serupa dilakukan untuk semua pohon lainnya. Tabel berikut menyajikan hasil prediksi dari semua model terhadap sampel ke-1.

Tabel 3. Prediksi seluruh model untuk sampel ke-1 data *testing*.

Model ke-	Prediksi untuk sampel ke-1 data testing
1	-0.1535
2	-0.1111
3	-0.1626
4	-0.1064
5	-0.1299
...	...
95	-0.0578
96	-0.1573

97	-0.1129
98	-0.0806
99	-0.1928
100	-0.1966

Tabel diatas menunjukkan hasil prediksi dari 100 model pohon keputusan terhadap sampel data *testing* ke-1. Setelah seluruh nilai prediksi diperoleh, maka nilai akhir prediksi untuk sampel ke-1 dihitung sebagai berikut:

$$\hat{y}_{test} = \frac{1}{B_{trees}} \sum_{i=1}^{B_{trees}} \hat{y}^{(b)}$$

$$\hat{y}_{test} = \frac{1}{100} ((-0.1535) + (-0.1111) + (-0.1626) + (-0.1064) + (-0.1299) + (-0.1410) + (-0.1149) + (-0.1690) + (-0.1789) + (-0.1206) + (-0.1375) + (-0.1175) + (-0.1335) + (-0.1677) + (-0.1615) + (-0.1838) + (-0.1859) + (-0.0914) + (-0.1166) + (-0.1511) + (-0.1423) + (-0.1679) + (-0.1735) + (-0.1652) + (-0.0632) + (-0.1637) + (-0.2452) + (-0.1756) + (-0.1654) + (-0.1653) + (-0.1610) + (-0.1153) + (-0.1638) + (-0.1953) + (-0.1949) + (-0.0913) + (-0.1159) + (-0.1055) + (-0.0879) + (-0.1218) + (-0.1698) + (-0.1954) + (-0.1687) + (-0.1959) + (-0.1955) + (-0.1474) + (-0.1558) + (-0.1052) + (-0.2159) + (-0.1535) + (-0.1318) + (-0.1056) + (-0.1586) + (-0.1527) + (-0.0926) + (-0.1279) + (-0.1080) + (-0.1695) + (-0.0884) + (-0.1056) + (-0.1655) + (-0.1631) + (-0.2452) + (-0.1593) + (-0.1461) + (-0.0918) + (-0.1689) + (-0.1897) + (-0.1989) + (-0.0997) + (-0.1861) + (-0.1352) + (-0.2008) + (-0.2475) + (-0.2056) + (-0.0895) + (-0.2055) + (-0.1928) + (-0.0605) + (-0.1492) + (-0.1440) + (-0.1748) + (-0.0611) + (-0.1775) + (-0.1548) + (-0.1209) + (-0.1524) + (-0.1911) + (-0.1201) + (-0.1194) + (-0.0686) + (-0.0982) + (-0.1794) + (-0.1262) + (-0.0578) + (-0.1573) + (-0.1129) + (-0.0806) + (-0.1928) + (-0.1966))$$

$$= -0.147$$

Perhitungan diatas adalah proses mencari prediksi dari sampel ke-1 pada data *testing* dengan hasil prediksi TMA adalah -0.147. Proses serupa dilakukan untuk mendapatkan hasil prediksi dari masing-masing sampel. Berikut ini adalah hasil prediksi dari semua sampel data *testing* yang dapat dilihat pada tabel 5.

Tabel 4. Hasil prediksi data *testing*

Suhu	Kelembaban Udara	Curah Hujan	TMA Aktual	TMA Prediksi
27.121	86.414	12.523	-0.1330	-0.1475
26.735	86.866	17.065	0.1241	0.0589
26.843	84.571	8.246	-0.2896	-0.3031
27.606	85.806	13.086	0.1151	-0.1924
27.503	85.613	6.258	-0.5229	-0.3167
26.738	86.069	7.357	-0.1933	-0.2705
26.940	85.600	8.468	-0.2904	-0.2738
27.329	85.226	14.005	0.0217	-0.1004
26.735	86.581	20.800	0.0969	0.1173
26.694	87.161	9.300	-0.2819	-0.2107
26.900	88.226	9.647	-0.0602	-0.1687
26.748	84.355	6.909	-0.3031	-0.3037
28.145	84.710	2.546	-0.6168	-0.5071
27.467	85.500	6.300	-0.2920	-0.3139

Tabel diatas merupakan hasil prediksi terhadap data testing yang telah dilakukan menggunakan model *Random forest regression*. Model ini dibangun berdasarkan proses melatih model (*training*) yang dilakukan pada data *training* dengan pendekatan ensemble learning, dimana sejumlah pohon keputusan (*Decision Trees*) digabungkan untuk mendapatkan prediksi akhir.

5) *Evaluasi Model Data Testing*

Evaluasi performa model pada data *testing* menggunakan metrik R*MSE* dan R^2 yang mencerminkan kemampuan generalisasi model Berikut ini merupakan perbandingan nilai aktual dan prediksi serta perhitungan galat dapat dilihat pada tabel 6.

Tabel 5. Perbandingan nilai aktual dan prediksi serta perhitungan galat

TMA_Aktual (y_i)	TMA_Prediksi ($\hat{y}_i$)	Error ($y_i - \hat{y}_i$)	Error^2 $(y_i - \hat{y}_i)^2$	$(y_i - \bar{y})^2$
-0.1330	-0.1475	0.0145	0.0002	0.0030
0.1241	0.0589	0.0652	0.0043	0.0971
-0.2896	-0.3031	0.0135	0.0002	0.0104

0.1151	-0.1924	0.3075	0.0945	0.0916
-0.5229	-0.3167	-0.2062	0.0425	0.1125
-0.1933	-0.2705	0.0773	0.0060	0.0000
-0.2904	-0.2738	-0.0165	0.0003	0.0106
0.0217	-0.1004	0.1221	0.0149	0.0438
0.0969	0.1173	-0.0204	0.0004	0.0809
-0.2819	-0.2107	-0.0712	0.0051	0.0089
-0.0602	-0.1687	0.1084	0.0118	0.0162
-0.3031	-0.3037	0.0006	0.0000	0.0134
-0.6168	-0.5071	-0.1098	0.0120	0.1843
-0.2920	-0.3139	0.0218	0.0005	0.0109
TOTAL		0.306763	0.19261	0.683513

Perhitungan evaluasi model pada data *testing* adalah sebagai berikut:
Root *Mean Squared Error* (R*MSE*)

$$RMSE = \sqrt{\frac{1}{14}}((0.0145)^2 + (0.0652)^2 + (0.0135)^2 + (0.3075)^2 + (-0.2062)^2 + (0.0773)^2 + (-0.0165)^2 + (0.1221)^2 + (-0.0204)^2 + (-0.0712)^2 + (0.1084)^2 + (0.0006)^2 + (-0.1098)^2 + (0.0218)^2))$$
$$= \sqrt{\frac{0.1926}{14}}$$
$$= \sqrt{0.0138}$$
$$= 0.1173$$

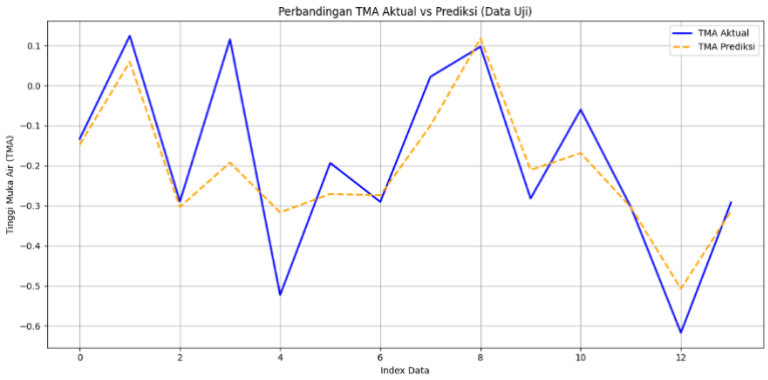
Koefisien Determinasi (R^2)

$$R^2 = 1 - \frac{0.1926}{0.6835}$$
$$= 1 - 0.2818$$
$$= 0.7182$$

Nilai R*MSE* pada data *testing* sebesar 0.1173 menunjukkan rata-rata kesalahan prediksi model dalam memprediksi TMA pada data yang belum pernah dilihat sebelumnya, dibandingkan dengan R*MSE* pada data *training* (0.0666), terdapat sedikit peningkatan kesalahan prediksi. Nilai R^2 pada data *testing* sebesar 0.7182 mengindikasikan bahwa sekitar 71.8% variasi dalam data *testing* dapat dijelaskan oleh model. Nilai ini lebih rendah dibandingkan dengan R^2 pada data *training* (0.9068), namun masih menunjukkan kemampuan prediksi yang cukup baik.

6) *Visualisasi Data Aktual dan Data Prediksi dari data testing*

Visualisasi perbandingan antara TMA *aktual* dan TMA prediksi pada data *testing* menunjukkan seberapa baik model *Random forest regression* mengikuti pola data sebenarnya. Grafik garis menunjukkan pergerakan TMA *aktual* dan TMA prediksi untuk setiap indeks data dalam set *testing*. Berikut adalah visualisasi perbandingan antara TMA *aktual* dan TMA prediksi pada data *testing* yang dapat dilihat pada gambar 3.



Gambar 2. Grafik nilai aktual dan prediksi data *testing*

Grafik diatas dapat diamati bahwa model *Random forest regression* cukup baik dalam mengikuti tren

pergerakan TMA, meskipun terdapat beberapa deviasi pada beberapa titik data. Model cenderung memprediksi nilai yang lebih rendah untuk TMA yang sangat tinggi dan nilai yang lebih tinggi untuk TMA yang sangat rendah, yang merupakan karakteristik umum dari model regresi.

Secara keseluruhan, pola yang konsisten antara TMA *aktual* dan TMA prediksi mengindikasikan bahwa model mampu menangkap hubungan antara variabel (suhu, kelembaban udara, dan curah hujan) dengan target (TMA). Kesesuaian pola ini juga tercermin dalam nilai R^2 yang cukup tinggi pada data *testing* (0.7182) yang menunjukkan bahwa model mampu menjelaskan sebagian besar variasi dalam data.

7) *Feature importance*

Feature importance dihitung berdasarkan seberapa besar kontribusi suatu fitur dalam mengurangi galat (Mean Squared Error) saat proses pemisahan (split) dilakukan. Hasil analisis feature importance untuk ketiga fitur merupakan akumulasi dari importance semua pohon keputusan. Sebagai contoh ilustrasi perhitungan Feature Importance (FI), dilakukan perhitungan pada node akar dari pohon keputusan ke-1. Pada node akar fitur yang digunakan untuk melakukan split adalah Curah Hujan, dengan jumlah sampel sebanyak 56. Sebelum dilakukan pemisahan, nilai MSE pada node ini adalah 0.046, total MSE setelah dilakukan split dari kedua cabang (anak kiri dan kanan) menjadi 0.018. Karena seluruh data pada node akar dipisahkan menggunakan fitur Curah Hujan, maka nilai $m_{t,f}$, yaitu jumlah sampel yang dipengaruhi oleh fitur ini yaitu 56. Perhitungan Fiture importance untuk fitur Curah Hujan adalah sebagai berikut:

$$FI(f) = \frac{m_{t,f}}{m}(MSE_{before} - MSE_{Total})$$
$$FI(3) = \frac{56}{56}(0.461 - 0.018) \ = 0.028$$

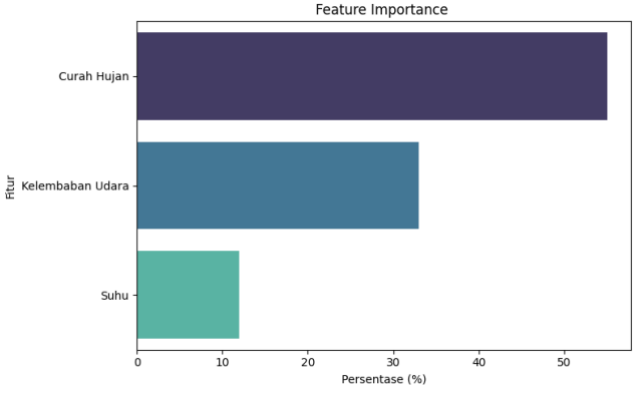
Nilai ini menunjukkan bahwa fitur Curah Hujan memberikan kontribusi sebesar 0.028 dalam mengurangi galat prediksi di node akar. Perhitungan ini hanya mencerminkan kontribusi di satu titik saja (node akar), sehingga untuk memperoleh nilai total Feature Importance dari sebuah fitur dalam satu pohon, diperlukan penjumlahan kontribusi dari semua node tempat fitur tersebut digunakan untuk melakukan split. Berikut ini merupakan akumulasi dari kontribusi masing-masing variabel pada model *Random forest regression* dapat dilihat pada tabel 7.

Tabel 6. *Feature Importance*

Fitur	Importance	Persentase
Curah Hujan	1.412268	55.048088
Kelembapan Udara	0.846384	32.990761
Suhu	0.306865	11.961152

Berdasarkan tabel diatas, dapat dilihat bahwa curah hujan memiliki kontribusi terbesar dalam memprediksi TMA dengan nilai importance yang sangat dominan yaitu 1.4122 (sekitar 55.05%). Kelembapan udara berada di urutan kedua dengan nilai importance 0.846 (sekitar 32.9%), dan suhu di urutan ketiga dengan nilai importance 0.3068 (sekitar 11.9%).

Hasil analisis ini menunjukkan bahwa curah hujan merupakan faktor yang paling berpengaruh terhadap Tinggi Muka Air (TMA) dengan kontribusi yang lebih dominan. Informasi feature importance ini sangat berharga untuk memahami faktor-faktor yang paling mempengaruhi TMA. Berikut ini adalah diagram batang yang menampilkan hasil persentase feature importance dapat dilihat pada gambar 4.



Gambar 3. Persentase Feature Importance

Gambar 8 menunjukkan persentase feature importance dalam model Random Forest Regression untuk memprediksi tinggi muka air. Hasil analisis mengindikasikan bahwa variabel curah hujan memiliki kontribusi terbesar, yaitu sebesar 55,05%, sehingga dikategorikan sebagai sangat penting karena berada di atas ambang batas 20%. Selanjutnya, variabel kelembaban udara berada di posisi kedua dengan kontribusi sebesar 32,9%, yang juga termasuk dalam kategori sangat penting. Sementara itu, variabel suhu memberikan kontribusi sebesar 11,9%, yang sesuai dengan kriteria berada pada rentang 10–20%, sehingga dikategorikan sebagai penting dalam memengaruhi prediksi tinggi muka air.

Kesimpulan

Berdasarkan hasil penelitian, dapat disimpulkan Model *Random Forest Regression* mampu memprediksi tinggi muka air (TMA) di lahan gambut Provinsi Jambi dengan performa yang baik dengan nilai RMSE = 0,1173 dan R^2 = 0,7182. Model ini dapat dimanfaatkan sebagai alat bantu dalam sistem peringatan dini dan perencanaan pengelolaan air secara berkelanjutan. Dari analisis feature importance, curah hujan terbukti menjadi faktor paling dominan yang memengaruhi TMA dengan kontribusi sebesar 55,05%, disusul kelembaban udara sebesar 32,9% dan suhu sebesar 11,9%. Hasil ini menegaskan peran penting variabel iklim, khususnya curah hujan, dalam pengendalian fluktuasi TMA pada ekosistem gambut.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan selama penyusunan penelitian ini. Ucapan khusus disampaikan kepada Badan Restorasi Gambut dan Mangrove (BRGM) atas kesempatan, data, dan arahan yang diberikan; kepada dosen pembimbing atas bimbingan dan motivasi; serta kepada keluarga dan rekan-rekan atas semangat dan dukungan moral. Penulis menyadari bahwa penelitian ini masih memiliki kekurangan, dan dengan terbuka menerima kritik serta saran yang membangun.

Daftar Pustaka

[1] Erianto Indra Putra and Idelia Lathiifah Puspadewi1, "PENGARUH KELEMBAPAN, SUHU UDARA DAN CURAH HUJAN TERHADAP KEJADIAN KEBAKARAN GAMBUT DI KABUPATEN TANJUNG JABUNG TIMUR, JAMBI," 2020.

[2] E. Suwananda, A. Saad, and A. Aswandi, "Pengaruh Curah Hujan terhadap Tinggi Muka Air Tanah Gambut Saat Fenomena El Nino di Sub Kesatuan Hidrologis Gambut Sungai Mendahara-Sungai Lagan," *Jurnal Ilmiah Universitas Batanghari Jambi*, vol. 24, no. 3, p. 2418, Oct. 2024, doi: 10.33087/jiubj.v24i3.5531.

[3] E. I. Putra and I. L. Puspadewi, "Pengaruh Kelembapan, Suhu Udara dan Curah Hujan terhadap Kejadian Kebakaran Gambut di Kabupaten Tanjung Jabung Timur, Jambi," *Journal of Tropical Silviculture*, vol. 11, no. 3, pp. 189–193, Dec. 2020, doi: 10.29244/j-siltrop.11.3.189-193.

[4] A. Firmansyah, M. F. Syahidin, and Y. S. Triana, "Prediksi Kebakaran Hutan Berdasarkan Titik Panas dan Iklim Menggunakan Algoritma Random Forest," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 10, no. 2, pp. 145–155, Sep. 2024, doi: 10.25077/TEKNOSI.v10i2.2024.145-155.

[5] T. Mahesti, K. D. Hartomo, and S. Y. J. Prasetyo, "Penerapan Algoritma Random Forest dalam Menganalisa Perubahan Suhu Permukaan Wilayah Kota Salatiga," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 4, p. 2074, Oct. 2022, doi: 10.30865/mib.v6i4.4603.

[6] A. Firmansyah, M. F. Syahidin, and Y. S. Triana, "Prediksi Kebakaran Hutan Berdasarkan Titik Panas dan Iklim Menggunakan Algoritma Random Forest," Sep. 2024, doi: 10.25077/TEKNOSI.v10i2.2024.145-155.

[7] M. Bagas, A. Darmawan, F. Dewanta, and S. Astuti, "Analisis Perbandingan Algoritma Decision Tree, Random Forest, dan Naïve Bayes untuk Prediksi Banjir di Desa Dayeuhkolot Comparative Analysis of Decision Tree, Random Forest, and Naïve Bayes Algorithm for Flood Prediction at Dayeuhkolot Village," *TELKA*, vol. 9, no. 1, pp. 52–61, 2023.

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

[9] S. Mahmuda, "Implementasi Metode Random Forest pada Kategori Konten Kanal Youtube," *Jurnal Jendela Matematika*, vol. 2, 2024.