JAMBURA JOURNAL OF INFORMATICS
Volume 7, No. 2 | April 2025: 1-10
ISSN 2685-4244, doi: 10.37905/jji.v1i1.30926

Prediksi kebakaran hutan ibu kota nusantara menggunakan produk MODIS dan algoritma machine learning

Syamsul Syahab Mangun, Kusrini

Magister Informatika, Universitas AMIKOM Yogyakarta, Indonesia

Riwayat Artikel:
Diterima 13 Maret 2025
Direvisi 23 April 2025
Disetujui 26 April 2025

Kata Kunci:
Algoritma Machine Learning
Ibu Kota Nusantara
Kebakaran hutan
MODIS
Prediksi

ABSTRACT. Forest fires in the Nusantara Capital City (IKN) area are a major threat to environmental sustainability and development. This study aims to build a forest fire prediction model using machine learning algorithms based on MODIS satellite data. The model was developed by comparing the performance of three regression algorithms: Linear Regression, Decision Tree Regressor (DTR), and Gradient Boosting Regressor (GBR). Data preprocessing involved data cleaning, feature selection, transformation, and initial visualization to understand data distribution patterns. The models were evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). The results showed that GBR achieved the best performance, with an accuracy of 98.83% and an R^2 score of 0.9883. The model proved effective in handling complex and non-linear spatiotemporal data. These findings support the development of an early warning system for forest fires in the IKN area and inform data-driven policy decisions by KLHK, BNPB, and the IKN Authority. Future research should consider integrating environmental variables and deep learning approaches.

ABSTRAK. Kebakaran hutan di wilayah Ibu Kota Nusantara (IKN) merupakan salah satu ancaman utama terhadap kelestarian lingkungan dan keberlangsungan pembangunan. Penelitian ini bertujuan membangun model prediksi kebakaran hutan menggunakan algoritma machine learning berbasis data satelit MODIS. Model dikembangkan dengan membandingkan kinerja tiga algoritma regresi, yaitu Regresi Linear, Decision Tree Regressor (DTR), dan Gradient Boosting Regressor (GBR). Tahap prapemrosesan data mencakup pembersihan data, seleksi fitur, transformasi, serta visualisasi awal untuk memahami pola distribusi data. Model dievaluasi menggunakan metrik Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan koefisien determinasi (R^2). Hasil menunjukkan bahwa algoritma GBR memberikan performa terbaik dengan akurasi mencapai 98,83% dan nilai R^2 sebesar 0,9883. Model ini dinilai efektif dalam menangani data spasial-temporal yang kompleks dan non-linear. Temuan ini dapat dimanfaatkan dalam pengembangan sistem peringatan dini kebakaran hutan di wilayah IKN dan mendukung pengambilan kebijakan oleh KLHK, BNPB, dan Otorita IKN. Penelitian selanjutnya disarankan mempertimbangkan variabel lingkungan dan pendekatan deep learning.

This is an open-access article under the CC–BY-SA license.

Penulis Korespondensi:
Syamsul Syahab Mangun,
Universitas Amikom Yogyakarta,
Jl. Ringroad Utara, Condong Catur, Depok, Sleman, D.I.Yogyakarta .
Email: syamsulsyahab@students.amikom.ac.id

PENDAHULUAN

Pembangunan Ibu Kota Nusantara (IKN) sebagai pusat pemerintahan dan ekonomi baru Indonesia menghadapi tantangan serius berupa kebakaran hutan yang mengancam kelestarian lingkungan dan keberlanjutan pembangunan (Wahyu & Frinaldi, 2024). Kebakaran hutan di wilayah tropis seperti Indonesia memberikan dampak multidimensi, mulai dari kerusakan ekologis, penurunan keanekaragaman hayati, gangguan iklim, hingga kerugian ekonomi yang signifikan (Najicha dkk., 2021). Berdasarkan laporan resmi Kementerian Lingkungan Hidup dan Kehutanan (KLHK), luas area

hutan dan lahan terdampak kebakaran di Indonesia pada periode 2015–2019 mencapai 1,6 juta hektar (Husen dkk., 2022). Fakta ini menegaskan pentingnya sistem deteksi dan prediksi dini yang akurat untuk mengantisipasi serta memitigasi dampak bencana kebakaran di wilayah strategis seperti IKN.

Kemajuan teknologi penginderaan jauh dan machine learning telah memberikan peluang baru dalam mengembangkan sistem prediksi kebakaran yang lebih adaptif dan akurat. Data satelit MODIS terbukti memainkan peran penting dalam pemantauan kebakaran hutan tropis (Shen Ruiyun, 2022), sementara algoritma machine learning seperti Random Forest (Salman dkk., 2024) dan Neural Networks (Kiranyaz dkk., 2021) telah menunjukkan efektivitas dalam pemodelan prediktif. Namun, beberapa studi terdahulu masih memiliki keterbatasan, antara lain kurang diterapkannya pendekatan ini pada ekosistem tropis spesifik seperti Kalimantan Timur, belum adanya evaluasi komprehensif antar algoritma di bawah kondisi iklim yang beragam, serta belum optimalnya integrasi hasil prediksi ke dalam sistem peringatan dini yang berbasis pengambilan keputusan.

Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model prediksi kebakaran hutan berbasis algoritma machine learning menggunakan data MODIS, dengan membandingkan performa Regresi Linear, Gradient Boosting Regressor (GBR), dan Decision Tree Regressor (DTR) di wilayah IKN. Kontribusi utama dari studi ini adalah menyediakan pendekatan komparatif yang teruji secara kuantitatif untuk mengidentifikasi algoritma terbaik dalam memprediksi anomali termal dan suhu permukaan di wilayah rawan kebakaran. Selain itu, studi ini membuka peluang integrasi model prediksi ke dalam sistem pemantauan kebakaran nasional seperti SiPongi atau dashboard KLHK (Wicaksono dkk., 2024), serta menjadi dasar perumusan kebijakan mitigasi berbasis data oleh otoritas terkait.

METODE

Jenis dan Pendekatan Penelitian

Penelitian ini merupakan studi kuantitatif dengan pendekatan komparatif-prediktif, yang bertujuan untuk membangun dan mengevaluasi model prediksi kebakaran hutan menggunakan algoritma machine learning. Model dikembangkan dengan memanfaatkan data spasial-temporal dari citra satelit MODIS dan dibandingkan performanya menggunakan tiga algoritma regresi, yaitu *Regresi Linear, Gradient Boosting Regressor (GBR),* dan *Decision Tree Regressor (DTR*).

Prosedur Penelitian

Tahapan penelitian dilakukan secara sistematis mulai dari identifikasi masalah, pengumpulan data, prapemrosesan, pemodelan, evaluasi, hingga penyusunan kesimpulan. Diagram alir proses penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan penelitian

Perencanaan Penelitian

Tahap awal ini dimulai dengan identifikasi masalah dan studi pustaka untuk memahami konteks kebakaran hutan di sekitar wilayah IKN.

Pemrosesan Data

Data yang digunakan dalam penelitian ini diperoleh dari MODIS (*Moderate Resolution Imaging Spectroradiometer)* dengan cakupan spasial sekitar wilayah IKN selama periode tahun 2000 hingga

2023. Batas wilayah ditetapkan pada rentang 115º-118.5º Bujur Timur (BT) untuk aspek longitudinal dari -2.5º Lintang Selatan (LS) hingga 2.5º Lintang Utara (LU). Dataset berisi atribut penting seperti *brightness* yang merupakan nilai temperatur radiasi dari *hotspot* dengan menunjukkan intensitas panas dalam suhu *Kelvin.* Adapun atribut lainnya seperti *Confidance* yang mengukur tingkat kepercayan deteksi titik panas dengan rentang 0- 100.

Tahap pra pemrosesan data dalam penelitian ini mencakup beberapa langkah yaitu proses *cleaning* dilakukan dengan menghapus data duplikat serta menghilangkan entri yang tidak lengkap guna memastikan kualitas dan konsistensi data. Kemudian proses *feature selection* dilakukan untuk memilih atribut paling relevan terhadap intensitas kebakaran yang bertujuan mengurangi kompleksitas model sekaligus meningkatkan akurasi prediksi. Adapun data divisualisasikan guna menyajikan informasi secara visual sehingga lebih mudah dipahami. Terakhir, data dibagi menjadi 80% data latih dan 20% data uji melalui teknik *train-test split*, namun, karena tidak diterapkannya metode *cross-validation,* hal ini diakui sebagai salah satu keterbatasan dalam penelitian. Sehingga, penelitian ini hanya mengandalkan *scikit-learn* yang merupakan *library machine learning* yang sangat popular dalam bahasa pemrograman *Phyton* (Tran dkk., 2022*).*

Analisis Data

Tahap analisis data melibatkan penggunaan metrik evaluasi seperti Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan koefisien determinasi (R^2) untuk mengukur kinerja model secara kuantitatif. Tiga algoritma yang digunakan dipilih berdasarkan kemampuan masing-masing dalam menangani hubungan non-linear.

Kesimpulan

Penelitian ditutup dengan tahap penarikan kesimpulan yang merangkum temuan utama, keterbatasan, dan kontribusi penelitian secara keseluruhan. Tahapan-tahapan ini mencerminkan pendekatan metodologis yang sistematis dan terstruktur dalam menjawab permasalahan penelitian.

HASIL DAN DISKUSI

Hasil Pengumpulan Data

Data yang dipakai dalam penelitian ini adalah hasil pengumpulan data dari MODIS (*Moderate Resolution Imaging Spectroradiometer*). Data disajikan dalam format .csv dengan jumlah total yang diperoleh sebanyak 1.492.022 data dan terdiri dari 15 atribut. Atribut ini terdiri dari *latitude, longtitude, brightness, scan, track, acq_date, acq_time, satelite, instrument, confidence, version, brigth_31, frp, daylight,* dan *type.* Adapun cuplikan dari dataset tersebut ditunjukkan pada Tabel 1.

Tabel 1. Cuplikan dataset

Latitude	*Longitude*	*brightness*	*scan*	..	*Bright_31*	*frp*	*daynight*	*type*
-0.5149	116.1463	310.1	1.0	..	292.8	6.8	N	0
0.6878	117.5666	302.8	1.7	..	291.5	8.2	N	0
1.1518	117.6893	310.1	2.1	..	293.0	14.8	D	0
1.1518	117.6894	304.6	1.4	..	280.7	5.8	D	0
…	…	…	…	..	…	…	…	…
0.0211	116.8739	315.3	1.1	..	295.6	8.7	D	0
0.4808	116.0806	312.3	1.0	..	295.0	6.9	D	0
2.1509	117.4968	320.6	1.0	..	297.5	10.6	D	0
2.1333	117.0875	317.4	1.0	..	290.6	13.3	D	0

Tabel 1 menunjukkan cuplikan data pada penelitian. Adapun cuplikan data tersebut memiliki 2 atribut target yaitu *latitude* dan *longitude*. Atribut tersebut dilakukan filterisasi data awal yang hanya

mencakup kawasan wilayah provinsi Kalimantan Timur saja. Selanjutnya, penjelasan dari setiap atribut ditunjukkan pada Tabel 2.

Tabel 2. Deskripsi atribut

Atribut	Deskripsi Atribut
Latitude	Garis lintang dari piksel kebakaran yang terdeteksi oleh satelit (derajat)
Longitude	Garis bujur dari piksel yang terdeteksi oleh satelit (derajat)
Brightness	Suhu kecerahan piksel kebakaran (dalam K)
Scan	Luas piksel MODIS di permukaan Bumi (sepanjang pemindaian: ΔS)
Track	Luas piksel MODIS di permukaan Bumi (sepanjang pemindaian: ΔT)
Acq_date	Tanggal terdeteksinya kebakaran
Acq_time	Waktu terdeteksinya kebakaran
Satellite	Sataelit yang digunakan untuk mendeteksi kebakaran. Either Terra (T) atau Aqua (A)
Instrument	MODIS
Confidence	Tingkat Kepercayaan deteksi (rentang 0-100)
version	Edisi dari data MODIS yang digunakan
Bright_31	Suhu kecerahan Band 31 piksel (dalam K)
frp	Daya radiasi kebakaran (dalam MW – megawatt)
daynight	Terdeteksi pada siang hari atau malam hari. Either Day (D) atau Night (N)
type	Jenis objek yang diamati oleh data MODIS

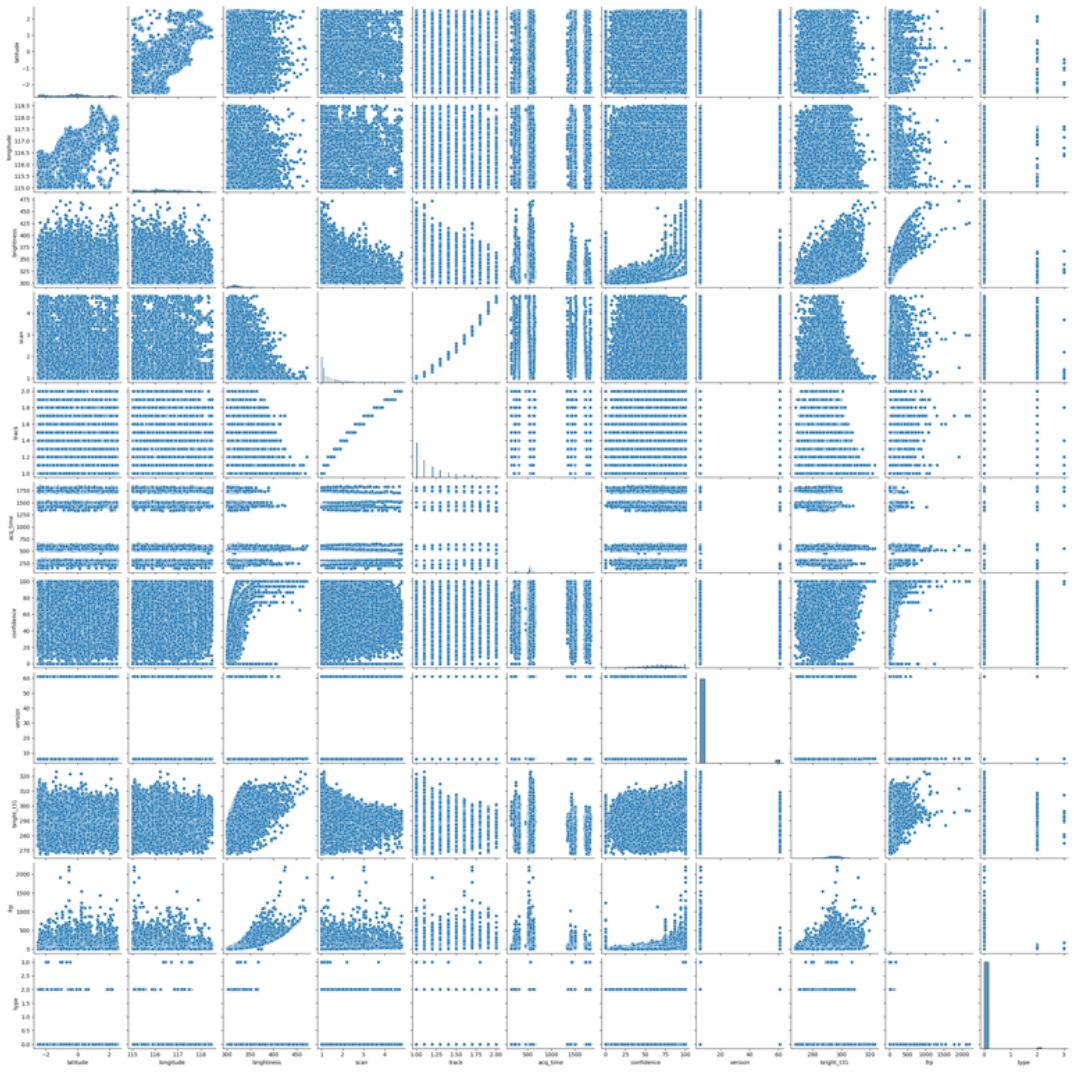
Hasil Pra-processing Data

Tahap pra-pemrosesan data merupakan bagian krusial dalam proses analisis, yang bertujuan memastikan bahwa data mentah yang telah dikumpulkan memiliki kualitas dan format yang layak untuk dianalisis lebih lanjut. Pada tahap ini dilakukan pemeriksaan terhadap kelengkapan dan konsistensi data guna mengidentifikasi serta menangani berbagai kesalahan, anomali, atau inkonsistensi yang berpotensi memengaruhi validitas hasil analisis. Langkah-langkah pra-pemrosesan meliputi pembersihan data (data cleaning), yaitu penghapusan nilai duplikat, data kosong, atau data yang tidak relevan. Selain itu, dilakukan penyaringan untuk mengurangi noise yang dapat mengganggu interpretasi model. Penyesuaian format data juga dilakukan, baik dari segi format file maupun satuan pengukuran, agar sesuai dengan kebutuhan teknis pemodelan machine learning. Sebagai bagian dari tahap eksplorasi awal, data divisualisasikan dalam berbagai bentuk seperti grafik, peta, atau diagram korelasi, dengan tujuan memberikan gambaran mengenai distribusi, pola, dan karakteristik data. Visualisasi ini membantu peneliti dalam memahami struktur data dan menentukan langkah-langkah analisis selanjutnya secara lebih terarah dan efisien.

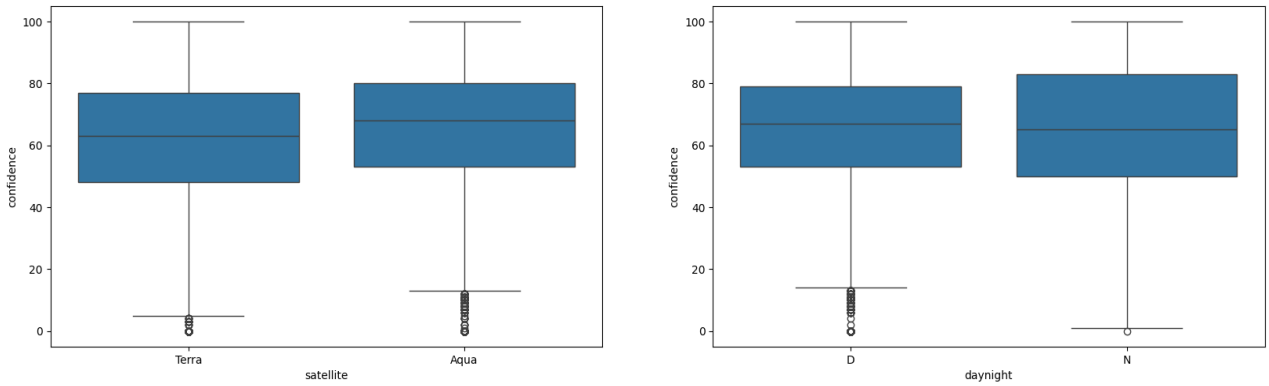
Pada Gambar 2 divisualisasikan variabel numerik menggunakan grafik pairplot, yang merupakan teknik visualisasi data efektif untuk menampilkan matriks *scatterplot*, memungkinkan peneliti melihat hubungan dan korelasi antara berbagai variabel numerik dalam dataset, seperti *latitude, longtitude, brightness, scan, track, acq_date, acq_time, satelite, instrument, confidence, version, brigth_31, frp, daylight,* dan *type*. Visualisasi ini membantu peneliti mengidentifikasi pola, tren, atau hubungan antar variabel, seperti korelasi antara *brightness* (kecerahan) dengan *frp* (*fire radiative power*), atau pengaruh *latitude dan longtitude* terhadap variabel lainnya.

Gambar 3 menampilkan dua *boxplot* yang membandingkan distribusi nilai *confidence* berdasarkan dua variabel kategorikal, yaitu satelit (*Terra* dan *Aqua*) serta waktu pengamatan (siang [D] dan malam [N]). Pada subplot pertama, distribusi *confidence* untuk kedua satelit menunjukkan pola yang hampir serupa, dengan nilai median berada di kisaran 65-70**.** Hal ini mengindikasikan bahwa baik satelit *Terra* maupun *Aqua* memiliki tingkat kepercayaan yang relatif sama dalam mendeteksi titik panas atau kebakaran hutan. Namun, terdapat beberapa *outlier* di bagian bawah, yang menandakan adanya deteksi dengan confidence rendah, yang mungkin terjadi akibat faktor atmosfer, gangguan sensor, atau kondisi lingkungan tertentu. Sementara itu, pada *subplot* kedua yang membandingkan *confidence* berdasarkan waktu pengamatan, terlihat bahwa distribusi *confidence* pada pengamatan

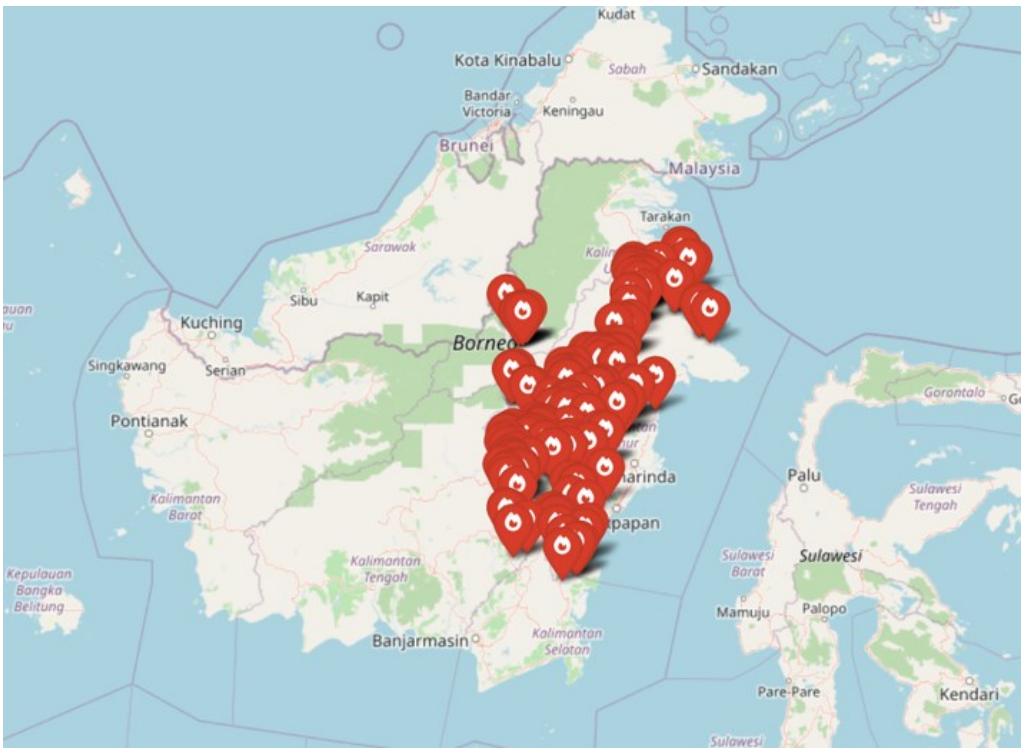
siang (D) dan malam (N) memiliki pola yang hampir mirip**,** dengan median yang tidak jauh berbeda. Namun, distribusi pada waktu malam (N) tampak lebih tersebar dibandingkan dengan siang (D), yang mengindikasikan adanya variasi yang lebih besar dalam tingkat kepercayaan deteksi kebakaran saat malam hari. Selain itu, jumlah outlier pada pengamatan malam lebih banyak, yang kemungkinan besar disebabkan oleh perbedaan kondisi pencahayaan dan tingkat *noise* yang lebih tinggi pada saat malam hari, sehingga mempengaruhi akurasi deteksi satelit.



Gambar 2. Grafik pairplot



Gambar 3. Grafik boxplot



Gambar 4. Peta interaktif

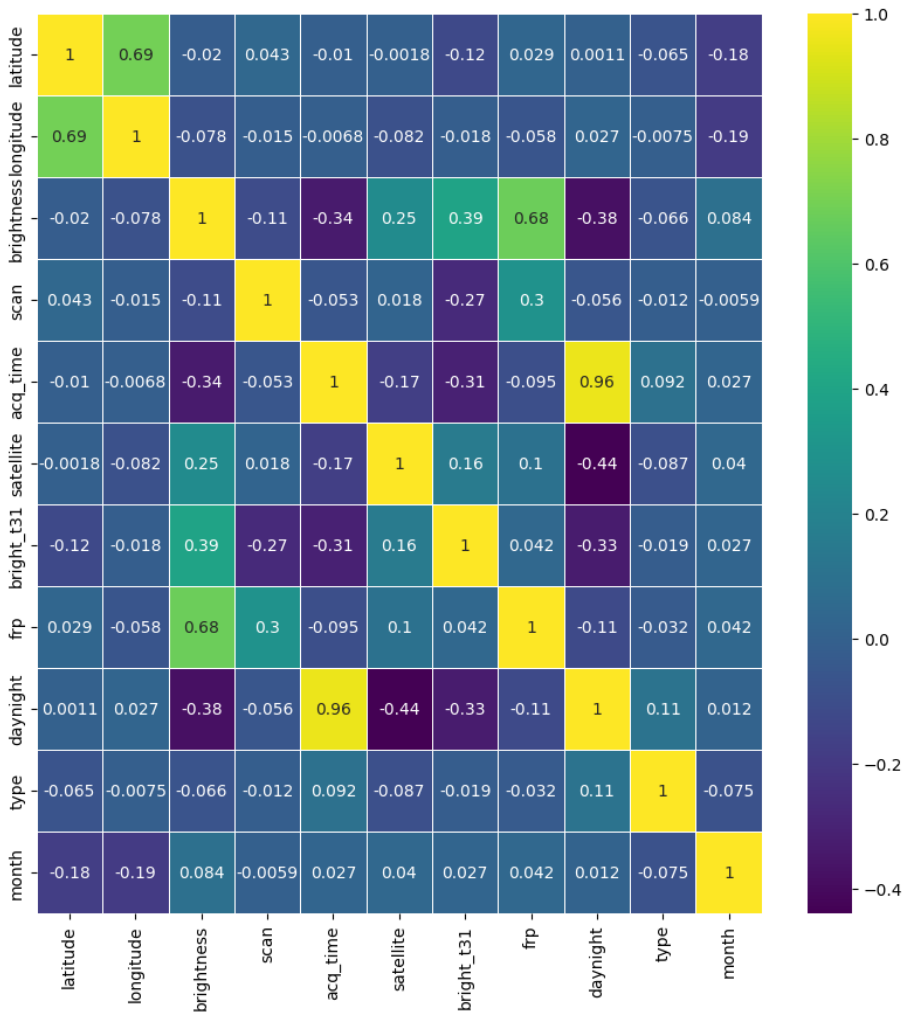
Gambar 4 menunjukkan hasil pemanfaatan *library pyton* yaitu *folium* untuk membuat peta interaktif (Nurpratama dkk., 2023), dimana di dalam folium pengguna bisa menggeser atau memperbesar memperkecil ukuran peta (Daffa dkk., 2022). Peta difokuskan pada wilayah Kalimantan Timur saja dengan menetapkan koordinat pusat tertentu serta dilengkapi dengan skala dan tingkat *zoom* untuk memberikan tampilan yang lebih jelas. Data yang digunakan kemudian disalin untuk menjaga integritas data asli sebelum diproses lebih lanjut. Dari data ini, ditambahkan penanda berbentuk api berwarna merah pada titik-titik kebakaran berdasarkan koordinatnya.

Hasil Persiapan Data

Pada tahap persiapan data, langkah penting yang dilakukan adalah menghilangkan data yang tidak relevan atau tidak mendukung tujuan penelitian, sehingga hanya informasi yang signifikan dan bermakna yang dipertahankan. Selain itu, data tambahan yang relevan dapat dimasukkan untuk memperkaya analisis dan memberikan pemahaman yang lebih mendalam serta kontekstual terhadap data utama yang sedang diteliti.

Proses yang dimaksud adalah menghapus kolom yang tidak diperlukan seperti *track*, *instrument*, dan *version*, serta melakukan encoding kategorikal pada kolom *satellite* dan *daynight* untuk mengubah nilai string menjadi numerik. Selanjutnya, kolom *month* ditambahkan dengan mengekstrak informasi bulan dari *acq_date*, dan dataset diambil sampelnya sebanyak 20% untuk mengurangi ukuran data. Setelah itu, variabel target *confidence* dipisahkan dari fitur, dan visualisasi *heatmap* digunakan untuk menganalisis korelasi antar variabel.

Gambar 5 merupakan *heatmap* korelasi yang menunjukkan hubungan antara variabel dalam dataset kebakaran hutan. Warna terang (kuning) menunjukkan korelasi positif yang kuat, sedangkan warna gelap (ungu/biru) menunjukkan korelasi negatif atau hubungan yang lemah. Beberapa korelasi yang menonjol adalah antara *latitude* dan *longitude* (0.69), *brightness* dan *frp* (0.68), serta *acquisition time* dan *day/night* (0.96). Korelasi negatif antara *day/night* dan *satellite* (-0.44) menunjukkan perbedaan deteksi berdasarkan waktu.



Gambar 5. Grafik barplot

Hasil Pembagian Data

Pada tahap ini ditentukan dataset terlebih dahulu di mana variabel prediktor dan variabel target dipisahkan, kemudian dataset dibagi menjadi data latih dan data uji. Pertama, variabel prediktor dipilih dari data frame dengan menggunakan kolom-kolom yang relevan, seperti *latitude, longitude, month, brightness, scan, acq_time, bright_t31, dan daynight*. Kolom-kolom ini dipilih karena dianggap sebagai faktor yang mempengaruhi atau berkontribusi terhadap variabel target, yaitu *Fire Radiative Power* (*frp*), yang merupakan energi yang dilepaskan oleh kebakaran. Variabel target didefinisikan sebagai kolom *frp*, yang merupakan nilai yang ingin diprediksi oleh model.

Selanjutnya, dataset dibagi menjadi dua subset menggunakan fungsi yang membagi dataset secara acak menjadi data latih dan data uji dengan proporsi 70% untuk data latih dan 30% untuk data uji. Data latih digunakan untuk melatih model, sementara data uji digunakan untuk menguji performa model dan memastikan bahwa model dapat generalisasi dengan baik pada data yang belum pernah dilihat sebelumnya. Pembagian dataset ini sangat penting untuk menghindari *overfitting*.

Hasil Evaluasi Matriks dan Analisa Prediksi

Tahap berikutnya melibatkan penggabungan beberapa algoritma regresi, yaitu regresi linear, *Gradient Boosting Regressor* (GBR), dan *Decision Tree Regression* (DTR), untuk membangun model prediksi. Kinerja model dievaluasi menggunakan metrik seperti *Mean Squared Error* (MSE),

Root Mean Squared Error (RMSE), *Mean Absolute Error* (MAE), *R-squared* (R^2), R-squared matriks dan tingkat akurasi, sebagaimana ditampilkan dalam gambar. Evaluasi ini bertujuan untuk menilai seberapa baik model dalam memprediksi nilai target yang sebenarnya. Hasil lengkap dari evaluasi kinerja model dapat dilihat dalam tabel yang menyajikan perbandingan berbagai metrik tersebut.

Tabel 3. Hasil evaluasi matriks

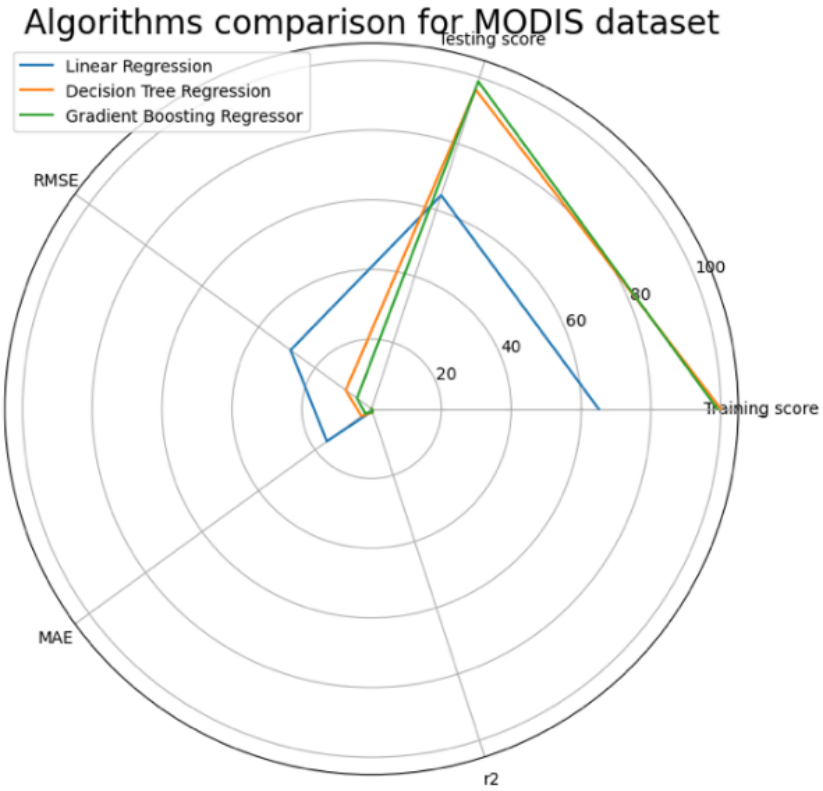
Metode	MSE	RMSE	MAE	R^2 *Score*	*R^2 metric*	Akurasi
Regresi Linear	822,89756	28,6861	15,8531	0,6446	0,65127	64.463 %
GBR	25,9135	5,1878	2,2472	0,9883	0,98817	98.837 %
DTR	83,6601	9,1465	3,5841	0,9628	1,00000	96.387 %

Tabel 3 adalah hasil evaluasi yang menunjukkan bahwa dari tiga model regresi yang diterapkan regresi linear, *Gradient Boosting Regressor* (GBR), dan *Decision Tree Regression* (DTR), algoritma GBR menunjukkan performa paling optimal dalam memprediksi kejadian kebakaran hutan di sekitar wilayah IKN berdasarkan data MODIS. GBR mencatat nilai *Mean Squared Error* (MSE) terendah sebesar 25,91 dan nilai koefisien determinasi (R^2) tertinggi sebesar 0,9883, yang berarti model ini mampu menjelaskan sekitar 98,83% variasi data target, dengan tingkat akurasi mencapai 98,84%. Sementara itu, DTR menghasilkan akurasi 96,39% dengan R^2 sebesar 0,9628, dan regresi linear hanya mencapai akurasi 64,46% dengan R^2 sebesar 0,6446. Hal ini menunjukkan bahwa GBR unggul dalam hal akurasi prediksi maupun kemampuan representasi pola data yang kompleks.

Keunggulan GBR terletak pada kemampuannya menangani hubungan non-linear dan mengurangi *overfitting* melalui teknik *boosting* yang menggabungkan banyak pohon keputusan lemah secara bertahap (Saidani dkk., 2022). Di sisi lain, meskipun DTR menunjukkan performa tinggi dalam data pelatihan, adanya perbedaan signifikan antara nilai *training* dan *testing score* menunjukkan kemungkinan *overfitting*, yang dapat dikurangi melalui teknik pruning atau penerapan ensemble seperti *Random Forest (*Salam dkk., 2021*)*.

Regresi linear, sebagai model dasar, memiliki keterbatasan dalam menangani data yang tidak berdistribusi linear seperti data MODIS yang dipengaruhi oleh variabel spasial-temporal serta faktor atmosfer seperti tutupan awan. Aspek ini juga menunjukkan potensi bias yang dapat memengaruhi hasil prediksi jika tidak ditangani melalui praprocessing data atau integrasi data multi-sumber. Akhirnya dalam memprediksi kebakaran hutan menggunakan data MODIS, peneliti membuat diagram radar yang membandingkan kinerja tiga algoritma tersebut. Dimana diagram ini memvisualisasikan lima metrik evaluasi, yaitu *training score, testing score, RMSE (Root Mean Squared Error), MAE (Mean Absolute Error),* dan *R^2 Score.*

Berdasarkan Gambar 6 terlihat bahwa garis berwarna *orange* yaitu *Decision Tree Regression* dan garis berwarna hijau sebagai *Gradient Boosting Regressor* memiliki *training score* yang sangat tinggi, mendekati angka 100, yang menunjukkan bahwa model tersebut sangat baik dalam menyesuaikan data pelatihan. Namun, perbedaan antara *training score* dan *testing score* pada DTR mengindikasikan kemungkinan *overfitting (López dkk., 2022)*, di mana model terlalu menyesuaikan data pelatihan tetapi kurang generalisasi terhadap data baru. Sementara itu, regresi linear memiliki *training* dan *testing score* yang lebih rendah dibandingkan dua model lainnya, yang mengindikasikan bahwa model ini mungkin memiliki performa yang kurang baik dalam menangkap pola kompleks dalam data MODIS. Pada metrik *error*, RMSE dan MAE regresi linear lebih tinggi dibandingkan DTR dan GBR, menunjukkan bahwa prediksinya kurang akurat. Sehingga dari ketiga model, GBR tampaknya memiliki keseimbangan terbaik antara *training score* dan *testing score*, dengan error yang lebih rendah dibandingkan DTR, sehingga lebih dapat diandalkan dalam melakukan prediksi kebakaran hutan berdasarkan data MODIS.



Gambar 6. Diagram radar

KESIMPULAN

Penelitian ini berhasil membangun model prediksi kebakaran hutan di wilayah Ibu Kota Nusantara (IKN) dengan memanfaatkan data citra satelit MODIS dan pendekatan machine learning. Dari tiga algoritma yang diuji—Regresi Linear, Decision Tree Regressor (DTR), dan Gradient Boosting Regressor (GBR)—model berbasis GBR menunjukkan performa terbaik, dengan nilai akurasi mencapai 98,83% dan R^2 sebesar 0,9883, menjadikannya model paling andal dalam mengidentifikasi pola kebakaran hutan tropis yang kompleks dan non-linear. Kontribusi utama dari penelitian ini terletak pada pengembangan pendekatan prediktif berbasis data spasial-temporal yang dapat mendukung sistem peringatan dini kebakaran hutan, khususnya di kawasan pembangunan strategis seperti IKN. Model prediksi yang dihasilkan juga memiliki potensi untuk diintegrasikan ke dalam sistem pemantauan nasional seperti Aplikasi SiPongi atau dashboard pengawasan KLHK, serta dapat dimanfaatkan oleh instansi seperti BNPB, KLHK, maupun otoritas IKN dalam pengambilan keputusan berbasis data. Meskipun demikian, penelitian ini memiliki sejumlah keterbatasan, seperti belum digunakannya variabel lingkungan tambahan seperti curah hujan, kelembaban tanah, atau tutupan vegetasi, serta tidak diterapkannya teknik validasi silang (cross-validation) dalam pengujian model. Oleh karena itu, penelitian lanjutan disarankan untuk mengeksplorasi penggunaan data satelit beresolusi lebih tinggi seperti VIIRS, serta menguji pendekatan deep learning untuk meningkatkan generalisasi dan ketepatan prediksi pada kondisi geografis dan iklim yang lebih beragam..

REFERENSI

Daffa, M., Satria Yudha Kartika, D., Permatasari, R., Pembangunan Nasional, U., Timur, J., Rungkut Madya, J., & Anyar, G. (2022). Penerapan folium untuk visualisasi hasil clustering persebaran tindakan kriminal di Kota Surabaya. *JUITIK*, *2*(3). https://journal.sinov.id/index.php/juitik/index

Halbouni, A., Gunawan, T. S., Habaebi, M. H., Halbouni, M., Kartiwi, M., & Ahmad, R. (2022). Machine Learning and Deep Learning Approaches for cybersecurity: A review dalam *IEEE Access* (Vol. 10,

hlm. 19572–19585). *Institute of Electrical and Electronics Engineers Inc.* https://doi.org/10.1109/ACCESS.2022.3151248

Husen, D., Sandi, D., & Bumbungan, S. (2022). Analisis Prediksi Kebakaran Hutan dengan Menggunakan Algoritma Random Forest Classifier. *16*(1). https://journal.uniku.ac.id/index.php/ilkom

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., & Inman, D. J. (2021). 1D convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing*, *151*. https://doi.org/10.1016/j.ymssp.2020.107398

López, O. A. M., Montesinos López, A., & Crossa, J. (2022). Multivariate statistical machine learning methods for genomic prediction. Dalam *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Springer International Publishing. https://doi.org/10.1007/978-3-030-89010-0

Najicha, F. U., Gusti, I., Ketut, A., & Handayani, R. (2021). Analysis of law enforcement and control effort about forest fire in west Kalimantan Province. https://doi.org/10.33172/jmb.v7i1.692

Saidani, O., Menzli, L. J., Ksibi, A., Alturki, N., & Alluhaidan, A. S. (2022). Predicting student employability through the internship context using gradient boosting models. *IEEE Access*, *10*, 46472–46489. https://doi.org/10.1109/ACCESS.2022.3170421

Salam, M. A., Azar, A. T., Elgendy, M. S., & Fouad, K. M. (2021). The effect of different dimensionality reduction techniques on machine learning overfitting problem. *International Journal of Advanced Computer Science and Applications*, *12*(4), 641–655. https://doi.org/10.14569/IJACSA.2021.0120480

Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random forest algorithm overview. *Babylonian Journal of Machine Learning*, *2024*, 69–79. https://doi.org/10.58496/bjml/2024/007

Shen Ruiyun. (2022). Application of MODIS data-based forest fire monitoring and assessment. Dalam *Highlights in Science, Engineering and Technology MSESCS* (Vol. 2022).

Tran, M. K., Panchal, S., Chauhan, V., Brahmbhatt, N., Mevawalla, A., Fraser, R., & Fowler, M. (2022). Python based scikit learn machine learning models for thermal and electrical performance prediction of high-capacity lithium-ion battery. *International Journal of Energy Research*, *46*(2), 786–794. https://doi.org/10.1002/er.7202

Wahyu, A., & Frinaldi, A. (2024). *Analisa dampak lingkungan ibu Kota Nusantara*. *01*(2), 29–33. https://jurnal.kopusindo.com/index.php/jkhkb

Wicaksono, A., Prihatin, P. S., Febrian, R. A., & Mulianto, B. (2024). Sipongi system: Navigating and fostering collaboration in Indonesia. *Journal of Infrastructure, Policy and Development*, *8*(3). https://doi.org/10.24294/jipd.v8i3.2875

Xiong, X., Angal, A., Chang, T., Chiang, K., Lei, N., Li, Y., Sun, J., Twedt, K., & Wu, A. (2020). Modis and viirs calibration and characterization in support of producing long term high-quality data products. *Remote Sensing*, *12*(19), 1–28. https://doi.org/10.3390/rs12193167

Nurpratama, Y. F., Dhian Satria Yudha Kartika, & Reisa Permatasari. (2023). Implementasi folium pada hasil klaster diabetes mellitus di Puskesmas Modopuro. *Jurnal Ilmiah Teknik Informatika dan Komunikasi*, *3*(3), 74–85. https://doi.org/10.55606/juitik.v3i3.623

Zheng, Y., Zhang, G., Tan, S., & Feng, L. (2023). Research on progress of forest fire monitoring with satellite remote sensing. *Agricultural & Rural Studies*, *1*(2), 0008. https://doi.org/10.59978/ar01020008

Zou, D., Yan, K., Pu, J., Gao, S., Li, W., Mu, X., Knyazikhin, Y., & Myneni, R. B. (2022). Revisit the performance of MODIS and VIIRS. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 8958–8973. https://doi.org/10.1109/JSTARS.2022.3214224