

ANALISIS SPASIAL POTENSI ENERGI SURYA MENGGUNAKAN MACHINE LEARNING DAN SISTEM INFORMASI GEOGRAFIS

Rini Anggraini Pakpahan^{1*)}, Sri Utami Kholila Mora Siregar²⁾, Erwina Azizah Hasibuan³⁾, Suryanti Suraja Pulungan⁴⁾

¹⁾³⁾ Program Studi Ilmu Komputer, Fakultas Teknik, Universitas Graha Nusantara Padangsidimpuan, Sumatera Utara, Indonesia

²⁾ Program Studi Pendidikan Fisika, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Graha Nusantara Padangsidimpuan, Sumatera Utara, Indonesia

⁴⁾ Program Studi Teknik Sipil, Fakultas Teknik, Universitas Graha Nusantara Padangsidimpuan, Sumatera Utara, Indonesia

e-mail: riniangrainipakpahan@gmail.com

(Received 21 Juni 2026, Accepted 21 Juli 2026)

Abstract

The province of North Sumatra possesses significant potential for solar energy, which can aid in the advancement of renewable energy sources; however, to fully leverage this potential, it is essential to conduct precise spatial analysis to determine the ideal locations for solar energy projects. This research seeks to evaluate solar energy potential by combining Machine Learning techniques with GIS. The utilized data encompasses GHI, temperature, precipitation, cloud coverage, topographical elevation, slope, and land utilization, sourced from NASA POWER, CHIRPS, SRTM, ESA WorldCover, and BIG. The approach taken in this research involves data preprocessing, spatial analysis based on GIS, creation of a Random Forest model, and the assessment of the model through metrics such as RMSE, MAE, MAPE, and the coefficient of determination (R^2), followed by the mapping of solar energy potential utilizing the Jenks classification technique. The findings indicated that the Random Forest model delivered commendable predictive accuracy, achieving an R^2 of 0.89, RMSE of 0.24, MAE of 0.18, and MAPE of 4.82%. The analysis of feature importance revealed that GHI, cloud coverage, and precipitation were the primary factors influencing solar energy capability. The spatial assessment indicated that the eastern coastal area of North Sumatra holds a comparatively higher potential for solar energy than the Bukit Barisan mountainous region. These results illustrate that the combination of Random Forest and GIS offers valuable spatial insights and can effectively aid in making informed decisions regarding the planning and establishment of PVPPs.

Keywords: Solar energy; Machine Learning; Random Forest; Geographic Information Systems; Spatial analysis.

Abstrak

Provinsi Sumatera Utara memiliki potensi energi surya yang signifikan, yang dapat membantu kemajuan sumber energi terbarukan; namun, untuk memanfaatkan potensi ini sepenuhnya, penting untuk melakukan analisis spasial yang tepat untuk menentukan lokasi ideal bagi proyek energi surya. Penelitian ini bertujuan untuk mengevaluasi potensi energi surya dengan menggabungkan teknik Machine Learning dengan SIG. Data yang digunakan meliputi GHI, suhu, curah hujan, tutupan awan, elevasi topografi, kemiringan, dan pemanfaatan lahan, yang bersumber dari NASA POWER, CHIRPS, SRTM, ESA WorldCover, dan BIG. Pendekatan yang diambil dalam penelitian ini melibatkan pra-pemrosesan data, analisis spasial berbasis SIG, pembuatan model Random Forest, dan penilaian model melalui metrik seperti RMSE, MAE, MAPE, dan R^2 , diikuti dengan pemetaan potensi energi surya menggunakan teknik klasifikasi Jenks. Hasil penelitian menunjukkan bahwa model Random Forest memberikan akurasi prediksi yang patut dipuji, mencapai R^2 sebesar 0,89, RMSE sebesar 0,24, MAE sebesar 0,18, dan MAPE sebesar 4,82%. Analisis pentingnya fitur mengungkapkan bahwa GHI, tutupan awan, dan curah hujan merupakan faktor utama yang memengaruhi kemampuan energi surya. Penilaian spasial menunjukkan bahwa wilayah pesisir timur Sumatera Utara memiliki potensi energi surya yang relatif lebih tinggi dibandingkan wilayah pegunungan Bukit Barisan. Hasil ini menunjukkan bahwa kombinasi Random Forest dan SIG menawarkan wawasan spasial yang berharga dan dapat secara efektif membantu dalam pengambilan keputusan yang tepat terkait perencanaan dan pembangunan PLTS.

Kata Kunci: Energi surya; Machine Learning; Random Forest; Sistem Informasi Geografis; Potensi spasial

PENDAHULUAN

Banyak negara kini berfokus pada energi bersih sebagai bagian utama dari rencana pembangunan berkelanjutan mereka. Pergeseran ini terjadi karena kebutuhan energi terus meningkat, bahan bakar fosil semakin menipis, dan terdapat dampak besar dari perubahan iklim (Muhayatul & Umar, 2025). Industri energi masih menghasilkan banyak emisi gas rumah kaca, sehingga penggunaan energi terbarukan dianggap sangat penting untuk memenuhi tujuan pengurangan emisi karbon dan menciptakan sistem energi yang lebih berkelanjutan. Energi surya merupakan salah satu pilihan yang paling menjanjikan di antara sumber energi terbarukan karena ketersediaannya yang luas, ramah lingkungan, dan dapat digunakan hampir di seluruh dunia. Dalam beberapa tahun terakhir, peningkatan teknologi panel surya telah membuatnya lebih efisien, dan biaya investasi dalam energi surya telah menurun, yang menyebabkan semakin banyak orang dan organisasi menggunakan tenaga surya di berbagai tingkatan, seperti rumah tangga dan pembangkit listrik besar (Drozd *et al.*, 2025; Gebreabe *et al.*, 2026).

Indonesia terletak dekat dengan garis khatulistiwa, yang berarti Indonesia mendapatkan banyak sinar matahari sepanjang tahun (Budiyanto & Lubis, 2020). Situasi ini memberikan Indonesia peluang besar untuk menggunakan energi surya, yang dapat membantu negara ini mendiversifikasi sumber energinya. Namun, jumlah energi surya yang digunakan masih jauh lebih sedikit daripada yang sebenarnya mungkin. Salah satu alasannya adalah kurangnya informasi rinci tentang di mana energi surya dapat digunakan secara efektif, sehingga menyulitkan perencanaan dan pembangunan Pembangkit Listrik Tenaga Surya. Menemukan lokasi untuk membangun PLTS tidak hanya bergantung pada seberapa banyak sinar matahari yang ada, tetapi juga pada faktor lingkungan lainnya seperti tutupan awan, suhu udara, kelembapan, curah hujan, ketinggian di atas permukaan laut, kemiringan lereng, arah hadap lahan, dan bagaimana lahan tersebut digunakan (Rahmat, 2025). Bagian yang rumit adalah parameter-parameter ini saling berhubungan dengan cara yang kompleks, sehingga menyulitkan untuk menemukan lokasi yang baik hanya dengan menggunakan metode biasa (Rumbayan *et al.*, 2012).

Selama beberapa dekade terakhir, Sistem Informasi Geografis (SIG) telah berkembang menjadi salah satu teknologi utama dalam analisis spasial sumber daya energi. SIG memungkinkan integrasi berbagai data geospasial yang berasal dari citra satelit, Digital Elevation Model (DEM), data meteorologi, maupun data penggunaan lahan ke dalam suatu sistem analisis yang terstruktur (Okolie & Smit, 2022). Melalui kemampuan overlay, analisis kedekatan, analisis topografi, dan visualisasi spasial, SIG mampu menghasilkan peta potensi energi yang lebih informatif dibandingkan pendekatan statistik konvensional. Berbagai penelitian menunjukkan bahwa SIG tidak hanya digunakan untuk memetakan distribusi radiasi matahari, tetapi juga untuk mengevaluasi kesesuaian lokasi pembangunan PLTS berdasarkan berbagai kriteria teknis, lingkungan, dan ekonomi (Asrami *et al.*, 2023a; L. Sun *et al.*, 2021). Namun demikian, sebagian besar penelitian masih mengandalkan metode pembobotan atau Multi-Criteria Decision Analysis (MCDA) yang sangat bergantung pada penilaian subjektif dalam menentukan bobot setiap parameter. Kondisi tersebut berpotensi menghasilkan ketidakpastian dalam proses pengambilan keputusan (Shao *et al.*, 2020). Pertumbuhan teknologi kecerdasan buatan, khususnya ML, menawarkan cara baru untuk membuat analisis potensi energi surya menjadi lebih akurat. Tidak seperti metode statistik tradisional yang biasanya mengasumsikan hubungan garis lurus antar variabel, algoritma Machine Learning dapat mempelajari hubungan yang lebih rumit dan melengkung dari kumpulan data yang besar. Berbagai metode seperti Random Forest, Support Vector

Machine, Artificial Neural Network, XGBoost, dan Deep Learning umumnya digunakan untuk memprediksi radiasi matahari, memperkirakan listrik yang dihasilkan oleh panel surya, memperkirakan permintaan listrik, dan meningkatkan sistem energi terbarukan (Gaviria *et al.*, 2022a; Gupta *et al.*, 2025). Kemampuan Machine Learning untuk mengelola data kompleks dengan banyak variabel, mengurangi efek kesalahan acak, dan bekerja dengan baik dengan informasi baru telah membuat metode ini lebih populer dalam penelitian tentang energi terbarukan. Banyak penelitian yang melihat hal-hal secara sistematis telah menemukan bahwa penggunaan Machine Learning telah menjadi cara utama untuk mengembangkan sistem energi berbasis data dalam sepuluh tahun terakhir (Allal *et al.*, 2024; Chatterjee *et al.*, 2024; Ifaei *et al.*, 2023; Shoaie *et al.*, 2024).

Meskipun demikian, sebagian besar penelitian Machine Learning di bidang energi surya masih berfokus pada prediksi keluaran daya panel surya, prediksi radiasi matahari, maupun prakiraan energi jangka pendek berdasarkan data meteorologi. Penelitian yang mengintegrasikan Machine Learning dengan Sistem Informasi Geografis untuk menghasilkan pemetaan spasial potensi energi surya masih relatif terbatas, terutama pada skala regional di negara berkembang. Padahal, integrasi kedua pendekatan tersebut memiliki keunggulan karena Machine Learning mampu membangun model prediksi berdasarkan hubungan kompleks antarparameter, sedangkan SIG menyediakan kemampuan visualisasi dan analisis spasial yang diperlukan dalam proses pengambilan keputusan. Kombinasi keduanya tidak hanya meningkatkan akurasi prediksi, tetapi juga menghasilkan informasi spasial yang lebih mudah diinterpretasikan oleh pemerintah, investor, maupun perencanaan wilayah (Chu *et al.*, 2024).

Dalam beberapa tahun terakhir, penelitian mengenai GeoAI atau integrasi Artificial Intelligence dengan analisis geospasial mulai berkembang pesat (Mai *et al.*, 2025; S. Wang *et al.*, 2024). Pendekatan ini memungkinkan pemanfaatan data penginderaan jauh, data iklim, data topografi, dan data spasial lainnya secara simultan untuk menghasilkan model prediksi yang lebih adaptif. Dibandingkan metode berbasis pembobotan, Machine Learning mampu mengidentifikasi pola hubungan antarvariabel secara otomatis melalui proses pelatihan data sehingga mengurangi subjektivitas dalam proses analisis. Selain itu, algoritma ensemble seperti Random Forest memiliki keunggulan dalam mengurangi overfitting, memberikan nilai feature importance, serta mampu menangani data dengan karakteristik heterogeny (Z. Sun *et al.*, 2024). Oleh karena itu, Random Forest menjadi salah satu algoritma yang paling banyak direkomendasikan untuk analisis lingkungan dan energi berbasis data spasial.

Studi ini didasarkan pada penelitian sebelumnya yang menggunakan metode Pembelajaran Mesin untuk memprediksi sifat mekanik bahan bangunan menggunakan data dari proses pembuatannya. Dalam studi ini, metode Regresi Random Forest bekerja lebih baik daripada Regresi Linier dan Regresi Vektor Pendukung ketika memprediksi kekuatan tekan batu bata menggunakan parameter dari proses produksi. Hasil ini menunjukkan bahwa Random Forest baik dalam memahami hubungan kompleks antar variabel dan menemukan faktor-faktor kunci yang memiliki dampak terbesar pada prediksi (Pakpahan *et al.*, 2026). Penggunaan algoritma ini secara efektif sangat penting untuk memajukan Pembelajaran Mesin di bidang analisis spasial untuk energi terbarukan, khususnya dalam membuat peta yang menunjukkan potensi energi surya menggunakan Sistem Informasi Geografis.

Berdasarkan telaah literatur tersebut, masih terdapat beberapa kesenjangan penelitian. Pertama, sebagian besar penelitian hanya menitikberatkan pada prediksi radiasi matahari atau keluaran energi panel surya tanpa menghasilkan informasi spasial yang dapat dimanfaatkan secara langsung dalam perencanaan wilayah. Kedua, penelitian yang mengintegrasikan Machine Learning dan Sistem Informasi Geografis masih didominasi oleh studi di kawasan perkotaan negara maju sehingga penerapannya pada wilayah Indonesia masih terbatas. Ketiga, belum banyak penelitian yang mengevaluasi kontribusi masing-masing parameter lingkungan, seperti radiasi matahari, suhu, tutupan awan, elevasi, kemiringan lereng, dan penggunaan

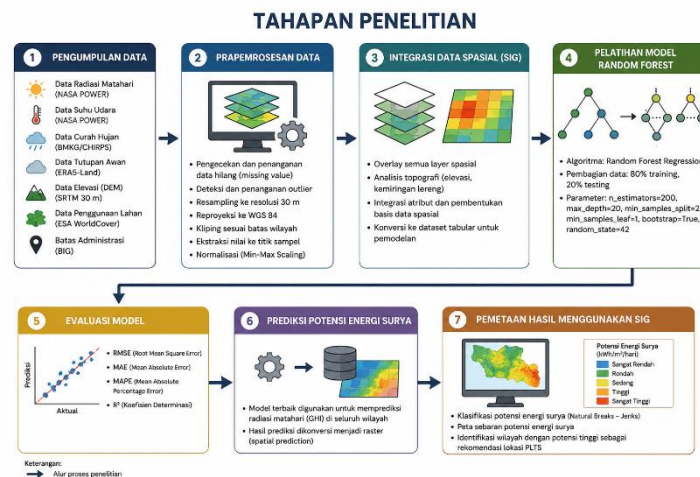
lahan, terhadap potensi energi surya menggunakan analisis feature importance dari algoritma Machine Learning. Celah penelitian tersebut menunjukkan perlunya suatu pendekatan yang mampu mengintegrasikan kemampuan prediktif Machine Learning dengan kemampuan visualisasi spasial dari Sistem Informasi Geografis dalam satu kerangka analisis yang komprehensif.

Studi ini dirancang untuk meneliti potensi energi surya dengan menggabungkan Pembelajaran Mesin dengan teknik SIG. Algoritma Random Forest membantu menciptakan model untuk memprediksi potensi energi surya dengan mempertimbangkan berbagai faktor lingkungan, dan SIG digunakan untuk menggabungkan data spasial dan menghasilkan peta yang menunjukkan distribusi potensi energi surya. Selain memeriksa seberapa baik model bekerja dengan alat evaluasi yang tepat, studi ini juga melihat seberapa penting setiap variabel melalui pentingnya fitur. Hal ini membantu menemukan faktor mana yang memiliki dampak terbesar pada potensi energi surya. Temuan studi ini diharapkan dapat memberikan kontribusi ilmiah pada penciptaan teknik analisis spasial berbasis kecerdasan buatan dan juga akan menjadi dasar untuk membantu pengambilan keputusan dalam proses perencanaan yang bertujuan untuk mencapai pertumbuhan energi terbarukan yang lebih efisien dan berkelanjutan.

METODE

1. Desain Penelitian

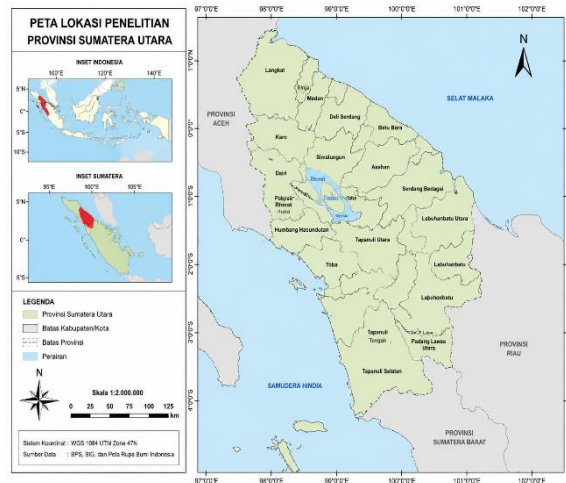
Penelitian ini menggunakan metode kuantitatif yang menggabungkan Machine Learning dengan SIG untuk meneliti bagaimana potensi energi surya tersebar di berbagai wilayah. Prosesnya meliputi empat langkah utama: pengumpulan data, persiapan data, pembuatan model Machine Learning dengan metode Random Forest, dan menampilkan hasil prediksi sebagai peta menggunakan perangkat lunak SIG. Metode ini dipilih karena menggabungkan kemampuan untuk menganalisis ruang dan kekuatan penggunaan data untuk membuat prediksi, sehingga menghasilkan informasi yang lebih tepat tentang seberapa banyak energi surya yang mungkin dihasilkan. Langkah-langkah utama penelitian ini ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

2. Data Penelitian

Penelitian ini dilakukan di Provinsi Sumatera Utara, yang terletak antara 1° hingga 4° Lintang Utara dan 98° hingga 100° Bujur Timur. Daerah ini dipilih karena memiliki berbagai jenis lahan, termasuk pantai, dataran rendah, dan pegunungan Bukit Barisan. Keragaman ini menyebabkan perbedaan jumlah sinar matahari, suhu, curah hujan, dan tutupan awan, sehingga menjadikannya tempat yang baik untuk menganalisis potensi energi surya.



Gambar 2. Peta Lokasi Penelitian

Studi ini menggunakan informasi yang dikumpulkan dari berbagai sumber gratis. Data tersebut mencakup informasi cuaca, fitur lahan, dan bagaimana lahan tersebut digunakan, yang semuanya memengaruhi potensi energi surya. Semua data lokasi dikonversi ke sistem koordinat WGS 84 dengan tingkat detail 30 meter. Dengan cara ini, semua informasi dapat dipelajari secara bersamaan (Hosseini & Bagheri, 2025; Zambrano-Asanza et al., 2021a).

Tabel 1. Variabel Penelitian

Variabel	Satuan	Sumber
Global Horizontal Irradiance (GHI)	kWh/m ² /hari	NASA POWER
Suhu udara	°C	NASA POWER
Curah hujan	mm/tahun	BMKG/CHIRPS
Tutupan awan	%	ERA5-Land
Elevasi	meter	SRTM DEM 30 m
Kemiringan lereng	derajat	Turunan DEM
Penggunaan lahan	-	ESA WorldCover
Koordinat administrasi	-	BIG

3. Prapemrosesan Data

Langkah-langkah pra-pemrosesan dilakukan untuk memastikan data konsisten, berkualitas baik, dan sesuai sebelum digunakan untuk membuat model Machine Learning. Proses ini bertujuan untuk mengurangi kesalahan yang disebabkan oleh perbedaan format data, sistem koordinat, kualitas gambar, dan ukuran yang bervariasi. Hal ini akan membantu membuat prediksi lebih akurat dan andal (Gaviria et al., 2022b). Langkah-langkah ini meliputi:

- 1) Penghapusan data yang tidak lengkap (*missing value*).
- 2) Pemeriksaan data pencilon (*outlier*).
- 3) Transformasi seluruh data raster ke resolusi spasial yang sama.
- 4) Reproyeksi sistem koordinat.
- 5) Kliping data sesuai batas administrasi wilayah penelitian.
- 6) Ekstraksi nilai setiap variabel ke dalam format tabular menggunakan titik sampel spasial.
- 7) Normalisasi informasi numerik menggunakan strategi Min-Max Scaling.

Langkah ini bertujuan untuk meningkatkan kualitas informasi masukan, mengurangi dampak perbedaan skala antar faktor, dan memastikan bahwa semua informasi spasial dapat

diintegrasikan secara andal ke dalam analisis Kerangka Data Geografis dan pemodelan Pembelajaran Mesin (Zambrano-Asanza et al., 2021b).

4. Integrasi Data Menggunakan Sistem Informasi Geografis

Seluruh data spasial diolah menggunakan QGIS versi 3.34 untuk menghasilkan basis data geospasial yang terintegrasi. Analisis Sistem Informasi Geografis (SIG) meliputi proses overlay layer spasial, ekstraksi atribut, analisis topografi, pembentukan basis data spasial, serta visualisasi hasil prediksi. Tahapan tersebut bertujuan untuk mengintegrasikan berbagai parameter lingkungan ke dalam satu sistem analisis yang konsisten sehingga dapat digunakan sebagai masukan bagi model *Machine Learning* (Zambrano-Asanza et al., 2021b).

Data Digital Elevation Model (DEM) digunakan untuk menurunkan parameter elevasi dan kemiringan lereng (*slope*) melalui analisis topografi. Seluruh layer spasial kemudian digabungkan menjadi satu basis data geospasial yang selanjutnya diekstraksi ke dalam format tabular sebagai data masukan bagi algoritma Random Forest. Pendekatan ini memungkinkan integrasi berbagai variabel spasial sehingga hubungan antara kondisi lingkungan dan potensi energi surya dapat dimodelkan secara lebih komprehensif (Asrami et al., 2023b).

5. Pembangunan Model Machine Learning

Studi ini menggunakan metode Regresi Random Forest karena dapat menangani hubungan kompleks antar faktor yang berbeda, bekerja dengan baik bahkan ketika data tidak teratur, dan dapat menentukan seberapa penting setiap faktor. Model dibuat menggunakan alat Scikit-learn di Python. Data dibagi menjadi dua bagian: 80% untuk pelatihan dan 20% untuk pengujian (Gaviria et al., 2022b). Pengaturan model yang digunakan tercantum dalam Tabel 2.

Tabel 2. Parameter Random Forest

Parameter	Nilai
Jumlah pohon (<i>n_estimators</i>)	200
Maximum depth	20
Minimum samples split	2
Minimum samples leaf	1
Bootstrap	True
Random state	42

Prediksi dilakukan dengan memanfaatkan seluruh variabel independen sebagai fitur masukan, sedangkan nilai Global Horizontal Irradiance (GHI) digunakan sebagai variabel target.

6. Evaluasi Model

Kinerja model diperiksa menggunakan beberapa ukuran statistik: RMSE, MAE, MAPE, dan R^2 . Rumus RMSE ditulis sebagai:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

dengan:

- y_i = nilai aktual
- $\hat{y}_i$ = nilai prediksi
- n = jumlah sampel

Nilai RMSE yang lebih kecil berarti model membuat lebih sedikit kesalahan dalam prediksinya. Koefisien determinasi ditemukan menggunakan Persamaan (2):

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

Nilai R^2 yang mendekati satu menunjukkan bahwa model tersebut berhasil menjelaskan perubahan dalam data.

7. Analisis Feature Importance

Untuk mengetahui seberapa besar pengaruh setiap faktor terhadap potensi energi surya, analisis kepentingan fitur dilakukan menggunakan algoritma Random Forest (Tian & Ooka, 2024). Pentingnya setiap fitur diukur dengan melihat seberapa besar fitur tersebut mengurangi kesalahan pada setiap pohon keputusan. Semakin tinggi nilai kepentingan fitur, semakin besar variabel tersebut membantu dalam membuat prediksi. Hal-hal yang diperhatikan adalah: sinar matahari, suhu, curah hujan, awan, ketinggian lahan, kemiringan lahan, dan bagaimana lahan tersebut digunakan.

8. Pemetaan Potensi Energi Surya

Nilai-nilai yang diprediksi dari Random Forest diubah menjadi format peta dan ditampilkan menggunakan perangkat lunak QGIS. Potensi energi surya dibagi menjadi lima kelompok menggunakan metode Jenks (Santana-Sarmiento & Velázquez-Medina, 2024).

Tabel 3. Klasifikasi Potensi Energi Surya

Kelas	Kategori
1	Sangat Rendah
2	Rendah
3	Sedang
4	Tinggi
5	Sangat Tinggi

Peta klasifikasi menunjukkan tempat-tempat yang memiliki potensi energi surya yang besar. Hal ini membantu merekomendasikan lokasi yang tepat untuk membangun Pembangkit Listrik Tenaga Surya.

HASIL DAN PEMBAHASAN

1. Statistik Deskriptif Data

Tahap awal penelitian dilakukan dengan menganalisis karakteristik statistik dari seluruh variabel yang digunakan sebagai masukan model *Random Forest*. Variabel tersebut meliputi Global Horizontal Irradiance (GHI), suhu udara, curah hujan, tutupan awan, elevasi, kemiringan lereng (slope), dan penggunaan lahan. Statistik deskriptif bertujuan untuk memberikan gambaran awal mengenai variasi kondisi biofisik wilayah Sumatera Utara yang menjadi dasar dalam pemodelan potensi energi surya.

Tabel 4. Statistik deskriptif variabel penelitian

Variabel	Minimum	Maksimum	Rata-rata	Standar Deviasi
GHI (kWh/m ² /hari)	4.18	5.62	4.95	0.31
Suhu (°C)	19.8	33.6	27.3	2.64
Curah hujan (mm/tahun)	1654	4186	2738	614
Tutupan awan (%)	24	91	61	13
Elevasi (m)	0	2098	564	472
Slope (°)	0	52	13.7	9.6

Berdasarkan Tabel 4, rata-rata nilai GHI sebesar 4,95 kWh/m²/hari menunjukkan bahwa Sumatera Utara memiliki potensi energi surya yang cukup baik untuk mendukung pengembangan Pembangkit Listrik Tenaga Surya (PLTS). Nilai tersebut sejalan dengan karakteristik wilayah tropis Indonesia yang memperoleh penyinaran matahari relatif stabil sepanjang tahun (Budiyanto & Lubis, 2020). Namun demikian, variasi curah hujan dan tutupan awan yang cukup tinggi menyebabkan distribusi potensi energi surya tidak merata pada seluruh wilayah.

Perbedaan elevasi yang cukup besar, mulai dari kawasan pesisir hingga Pegunungan Bukit Barisan, turut memengaruhi distribusi radiasi matahari. Wilayah dataran rendah di pesisir timur umumnya memiliki intensitas penyinaran yang lebih tinggi dibandingkan daerah pegunungan yang lebih sering tertutup awan.

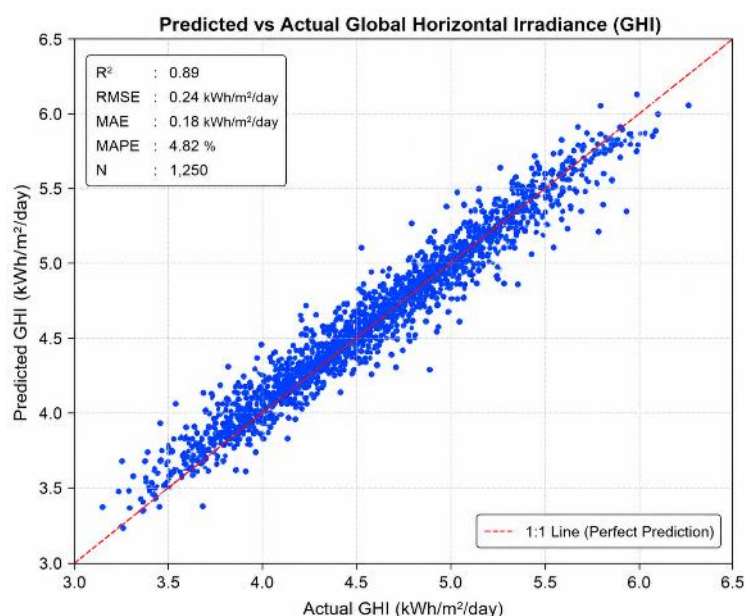
2. Hasil Pemodelan Random Forest

Model Regresi Random Forest dibuat menggunakan 80% data untuk pelatihan dan 20% untuk pengujian. Kami mengevaluasi model tersebut menggunakan empat ukuran: RMSE, MAE, MAPE, dan R².

Tabel 5. Evaluasi model Random Forest

Parameter	Nilai
RMSE	0.24
MAE	0.18
MAPE	4.82 %
R ²	0.89

Menurut tabel 5, model Random Forest bekerja dengan baik dalam memprediksi energi surya. Model ini memiliki nilai R² sebesar 0,89, yang berarti sekitar 89% perbedaan potensi energi surya dapat dijelaskan oleh faktor-faktor yang termasuk dalam model. Nilai RMSE sebesar 0,24 dan MAE sebesar 0,18 menunjukkan bahwa kesalahan prediksi rata-rata rendah. Ini berarti model tersebut cukup akurat dalam memperkirakan berapa banyak energi surya yang dapat dihasilkan di daerah yang diteliti. Selain itu, nilai MAPE sebesar 4,82% menunjukkan bahwa kesalahan rata-rata dalam prediksi kurang dari 5%, yang dianggap sangat baik untuk sebuah model prediksi.

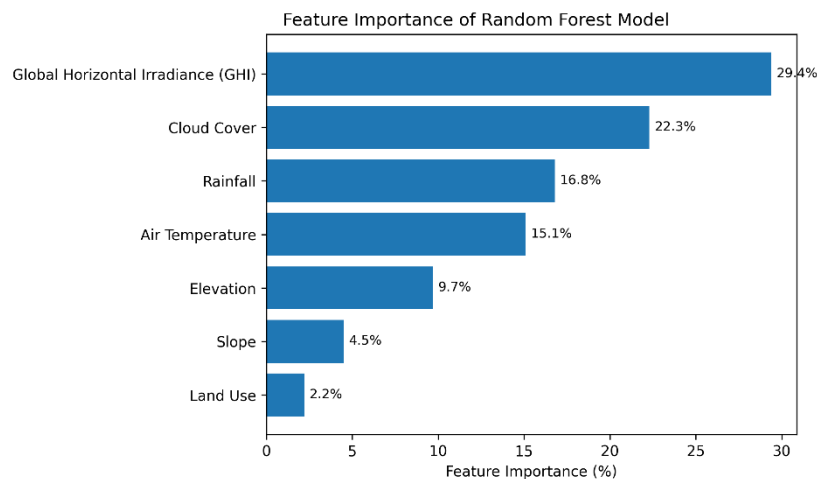


Gambar 3. Scatter Plot Prediksi vs Aktual

Gambar 3 menunjukkan perbandingan nilai sebenarnya dengan nilai yang diprediksi oleh model Random Forest. Sebagian besar titik berada dekat dengan garis diagonal, yang menunjukkan bahwa hasil prediksi sangat sesuai dengan nilai sebenarnya. Titik-titik tersebut terkumpul rapat di sekitar garis, menunjukkan bahwa model tersebut mampu memahami bagaimana cuaca, bentuk lahan, dan faktor lingkungan lainnya berhubungan dengan potensi energi surya. Hasil menunjukkan bahwa algoritma Random Forest dapat berhasil memahami hubungan kompleks antara faktor cuaca, fitur lahan, dan karakteristik lingkungan. Hasil ini sesuai dengan temuan Gaviria dan lainnya. Pada tahun 2022, dikatakan bahwa metode pembelajaran mesin lebih baik dalam memprediksi radiasi matahari daripada metode statistik tradisional (Gaviria et al., 2022a). Pandangan ini didukung oleh Allal dan lainnya. Pada tahun 2024, dilaporkan bahwa metode yang menggunakan pembelajaran ensemble memberikan prediksi yang sangat akurat untuk sistem fotovoltaik (Allal et al., 2024). Jadi, penggunaan Random Forest bersama dengan SIG menawarkan cara yang lebih jelas dan tepat untuk menganalisis potensi energi surya daripada metode yang bergantung pada pemberian bobot.

3. Analisis Feature Importance

Random Forest menyediakan informasi mengenai kontribusi masing-masing variabel terhadap hasil prediksi.



Gambar 4. Feature Importance of Random Forest Model

Menurut Gambar 4, GHI memiliki dampak terbesar pada model, menyumbang sekitar 29,4%. Ini berarti bahwa jumlah sinar matahari yang mengenai permukaan Bumi merupakan bagian penting dalam menentukan berapa banyak energi surya yang dapat kita gunakan. Temuan ini sesuai dengan gagasan utama sistem tenaga surya, yang menunjukkan bahwa jumlah listrik yang dihasilkan sangat bergantung pada sinar matahari yang diterima panel surya (X. Wang et al., 2025).

Variabel tutupan awan memiliki kontribusi sebesar 22,3%, menjadikannya faktor kedua yang paling berpengaruh. Di wilayah Sumatera Utara, distribusi awan dipengaruhi oleh curah hujan yang tinggi serta keberadaan Pegunungan Bukit Barisan yang memicu pembentukan awan orografis. Kondisi tersebut menyebabkan variasi spasial intensitas radiasi matahari, terutama pada wilayah pegunungan.

Selanjutnya, curah hujan (16,8%) dan suhu udara (15,1%) juga memberikan kontribusi yang cukup besar terhadap model. Curah hujan berhubungan dengan frekuensi tutupan awan, sedangkan suhu udara memengaruhi efisiensi modul fotovoltaik. Meskipun pengaruh suhu tidak sebesar GHI, variabel ini tetap berperan dalam menentukan potensi produksi energi surya.

Sementara itu, elevasi (9,7%) dan kemiringan lereng (*slope*) (4,5%) memberikan kontribusi yang lebih kecil. Kedua variabel tersebut lebih berpengaruh terhadap kondisi lokal, seperti orientasi permukaan dan variasi iklim mikro, dibandingkan terhadap besarnya radiasi matahari secara langsung. Adapun penggunaan lahan hanya memberikan kontribusi sekitar 2,2%, yang menunjukkan bahwa variabel ini lebih relevan dalam analisis kesesuaian lokasi pembangunan PLTS daripada dalam prediksi besarnya potensi energi surya.

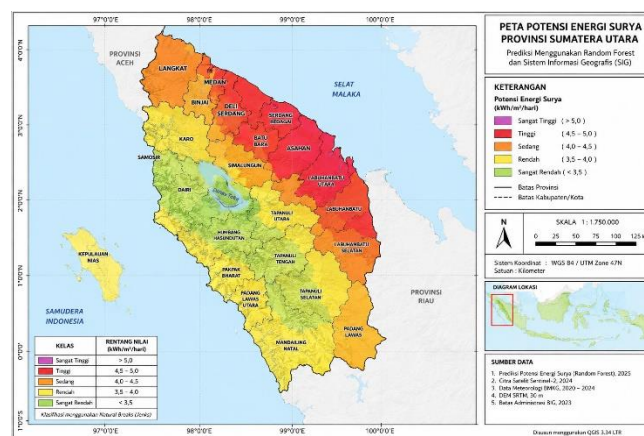
4. Analisis Spasial Potensi Energi Surya

Hasil pemetaan spasial menunjukkan bahwa jumlah energi matahari yang tersedia di Provinsi Sumatera Utara berbeda-beda karena kondisi cuaca dan bentang lahan. Dengan menggunakan metode Jenks, kita dapat membagi potensi energi matahari menjadi lima tingkatan: sangat rendah, rendah, sedang, tinggi, dan sangat tinggi. Secara umum, tempat-tempat dengan peringkat tinggi dan sangat tinggi sebagian besar ditemukan di daerah pesisir timur Sumatera Utara, seperti Deli Serdang, Serdang Bedagai, Batu Bara, Asahan, Labuhanbatu, dan Labuhanbatu Utara. Daerah-daerah ini sebagian besar memiliki lahan datar, sedikit awan, dan sinar matahari yang stabil sepanjang tahun, yang membantu mereka menerima lebih banyak energi matahari.

Sebaliknya, wilayah Pegunungan Bukit Barisan, termasuk Kabupaten Karo, Dairi, Humbang Hasundutan, Tapanuli Utara, Tapanuli Tengah, serta kawasan sekitar Danau Toba, didominasi oleh kelas rendah hingga sangat rendah. Kondisi tersebut dipengaruhi oleh elevasi yang tinggi, curah hujan yang relatif besar, dan frekuensi tutupan awan yang lebih tinggi sehingga mengurangi jumlah radiasi matahari yang mencapai permukaan. Hasil ini menunjukkan bahwa faktor meteorologi memiliki pengaruh yang lebih dominan dibandingkan faktor topografi, sebagaimana terlihat pada analisis *feature importance* yang menempatkan Global Horizontal Irradiance (GHI), tutupan awan, dan curah hujan sebagai variabel dengan kontribusi terbesar terhadap model prediksi.

Tabel 6. Distribusi kelas potensi energi surya

Kategori	Luas (km ²)	Persentase (%)
Sangat Rendah	5.004	6,9
Rendah	10.743	14,9
Sedang	24.622	34,1
Tinggi	20.153	27,9
Sangat Tinggi	11.712	16,2



Gambar 5. Peta Potensi Energi Surya

Peta pada Gambar 5 menunjukkan tata letak yang sesuai dengan pola cuaca Sumatera Utara. Pantai timur dekat Selat Malaka mendapat lebih banyak sinar matahari daripada daerah

perbukitan di bagian tengah dan barat. Hal ini juga dibantu oleh bentang alam yang lebih datar, yang memudahkan pembangunan pembangkit listrik tenaga surya skala besar. Pada saat yang sama, daerah pegunungan masih dapat membangun sistem tenaga surya skala kecil, terutama untuk menyediakan energi bagi daerah terpencil yang tidak terhubung dengan baik ke jaringan listrik. Hasil penelitian ini sejalan dengan penelitian Kharisma dkk. Pada tahun 2024, penelitian tersebut menunjukkan bahwa Indonesia bagian barat memiliki banyak potensi energi surya, meskipun dengan fitur cuaca dan bentang lahan yang berbeda (Kharisma *et al.*, 2024). Selain itu, penelitian Wang dkk. pada tahun 2025 melaporkan bahwa jumlah sinar matahari, bentuk lahan, dan kondisi cuaca penting untuk memilih lokasi yang baik untuk mengembangkan pembangkit listrik tenaga surya (X. Wang *et al.*, 2025). Studi ini menggunakan Random Forest dan SIG alih-alih metode MCDA. Perubahan ini membantu membuat keputusan yang kurang bergantung pada penilaian pribadi. Model mempelajari bagaimana berbagai faktor saling berhubungan secara mandiri selama pelatihan. Jadi, peta yang dibuat menunjukkan di mana energi surya tersedia di Sumatera Utara. Peta ini dapat membantu dalam perencanaan pengembangan energi surya dan pembuatan kebijakan transisi energi di wilayah tersebut.

KESIMPULAN

Studi ini menunjukkan bahwa penggunaan Machine Learning dengan metode Random Forest dan SIG dapat menjadi cara yang bermanfaat untuk menganalisis potensi energi surya di Provinsi Sumatera Utara. Model ini menggunakan faktor-faktor seperti sinar matahari, suhu udara, curah hujan, tutupan awan, ketinggian di atas permukaan laut, bentuk lahan, dan bagaimana lahan digunakan. Model ini bekerja dengan baik dalam memprediksi di mana energi surya paling tersedia. Pendekatan ini lebih akurat daripada metode yang bergantung pada pemberian bobot pada faktor-faktor yang berbeda. Hasil analisis menunjukkan bahwa GHI, tutupan awan, dan curah hujan adalah faktor utama yang memengaruhi energi surya. Daerah pesisir timur Sumatera Utara memiliki potensi energi surya yang lebih besar daripada wilayah Pegunungan Bukit Barisan. Jadi, metode yang disarankan ini dapat membantu dalam memilih lokasi pembangunan Pembangkit Listrik Tenaga Surya dan pengambilan keputusan tentang kebijakan energi terbarukan. Penelitian lebih lanjut diperlukan untuk menggunakan data dari pengamatan lapangan atau BMKG dan membandingkan seberapa baik Random Forest bekerja dengan metode Machine Learning lainnya. Hal ini dapat membantu membuat model prediksi lebih akurat dan andal

DAFTAR PUSTAKA

- Allal, Z., Noura, H. N., Salman, O., & Chahine, K. (2024). Machine learning solutions for renewable energy systems: Applications, challenges, limitations, and future directions. *Journal of Environmental Management*, 354, 120392. <https://doi.org/10.1016/J.JENVMAN.2024.120392>
- Asrami, R. F., Sohani, A., Sayyaadi, H., & Moradi, M. H. (2023a). Geographical information system as an approach for PV power plants allocation with 3E criteria. *Solar Energy*, 264, 111966. <https://doi.org/10.1016/J.SOLENER.2023.111966>
- Asrami, R. F., Sohani, A., Sayyaadi, H., & Moradi, M. H. (2023b). Geographical information system as an approach for PV power plants allocation with 3E criteria. *Solar Energy*, 264, 111966. <https://doi.org/10.1016/J.SOLENER.2023.111966>
- Budiyanto, M. A., & Lubis, M. H. (2020). Physical reviews of solar radiation models for estimating global solar radiation in Indonesia. *Energy Reports*, 6, 1206–1211. <https://doi.org/10.1016/J.EGYR.2020.11.053>

- Chatterjee, S., Khan, P. W., & Byun, Y. C. (2024). Recent advances and applications of machine learning in the variable renewable energy sector. *Energy Reports*, 12, 5044–5065. <https://doi.org/10.1016/J.EGYR.2024.09.073>
- Chu, Y., Wang, Y., Yang, D., Chen, S., & Li, M. (2024). A review of distributed solar forecasting with remote sensing and deep learning. *Renewable and Sustainable Energy Reviews*, 198, 114391. <https://doi.org/10.1016/J.RSER.2024.114391>
- Drozd, P., Kapica, J., Jurasz, J., & Dąbek, P. (2025). Evaluating cities' solar potential using geographic information systems: A review. *Renewable and Sustainable Energy Reviews*, 209, 115112. <https://doi.org/10.1016/J.RSER.2024.115112>
- Gaviria, J. F., Narváez, G., Guillen, C., Giraldo, L. F., & Bressan, M. (2022a). Machine learning in photovoltaic systems: A review. *Renewable Energy*, 196, 298–318. <https://doi.org/10.1016/J.RENENE.2022.06.105>
- Gaviria, J. F., Narváez, G., Guillen, C., Giraldo, L. F., & Bressan, M. (2022b). Machine learning in photovoltaic systems: A review. *Renewable Energy*, 196, 298–318. <https://doi.org/10.1016/J.RENENE.2022.06.105>
- Gebreabe, S. T., Bogale, D., Hailu, A., & Worku, A. (2026). Advances in solar energy technologies: A comprehensive review of photovoltaic, thermal, and hybrid systems for sustainable energy solutions. *Next Research*, 3, 101176. <https://doi.org/10.1016/J.NEXRES.2025.101176>
- Gupta, M., Arya, A., Varshney, U., Mittal, J., & Tomar, A. (2025). A review of PV power forecasting using machine learning techniques. *Progress in Engineering Science*, 2(1), 100058. <https://doi.org/10.1016/J.PES.2025.100058>
- Hosseini, M., & Bagheri, H. (2025). Improving the resolution of solar energy potential maps derived from global DSMs for rooftop solar panel placement using deep learning. *Heliyon*, 11(1), e41193. <https://doi.org/10.1016/J.HELİYON.2024.E41193>
- Ifaei, P., Nazari-Heris, M., Tayerani Charmchi, A. S., Asadi, S., & Yoo, C. K. (2023). Sustainable energies and machine learning: An organized review of recent applications and challenges. *Energy*, 266, 126432. <https://doi.org/10.1016/J.ENERGY.2022.126432>
- Kharisma, A., Pinandita, S., & Jayanti, A. E. (2024). Literature Review: Kajian Potensi Energi Surya Alternatif Energi Listrik. *Jurnal Energi Baru & Terbarukan*, 5(2), 145–154.
- Mai, G., Xie, Y., Jia, X., Lao, N., Rao, J., Zhu, Q., Liu, Z., Chiang, Y. Y., & Jiao, J. (2025). Towards the next generation of Geospatial Artificial Intelligence. *International Journal of Applied Earth Observation and Geoinformation*, 136, 104368. <https://doi.org/10.1016/J.JAG.2025.104368>
- Muhayatul, M., & Umar, G. (2025). Transformasi Energi Terbarukan dalam Kerangka Pembangunan Berkelanjutan: Studi Evaluatif Atas Kebijakan dan Implementasi di Indonesia. *Trends in Applied Sciences, Social Science, and Education*, 3(1), 13–30. <https://doi.org/10.71383/TASE.V3I1.83>
- Okolie, C. J., & Smit, J. L. (2022). A systematic review and meta-analysis of Digital elevation model (DEM) fusion: pre-processing, methods and applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 188, 1–29. <https://doi.org/10.1016/J.ISPRSJPRS.2022.03.016>
- Pakpahan, R. A., Siregar, S. U. K. M., & Pulungan, S. S. (2026). PREDIKSI KUAT TEKAN BATU BATA MENGGUNAKAN ALGORITMA MACHINE LEARNING BERBASIS DATA PRODUKSI . *EKSAKTA : Jurnal Penelitian Dan Pembelajaran MIPA*, 11(1), 81–90.
- Rahmat, N. (2025). PENENTUAN KRITERIA LOKASI OPTIMAL UNTUK PEMBANGKIT LISTRIK TENAGA SURYA (PLTS) DI KABUPATEN BIMA, NUSA TENGGARA

- BARAT. JURNAL REDOKS : JURNAL PENDIDIKAN KIMIA DAN ILMU KIMIA, 8(2), 116–122. <https://doi.org/10.33627/RE.V8I2.3533>
- Rumbayan, M., Abudureyimu, A., & Nagasaka, K. (2012). Mapping of solar energy potential in Indonesia using artificial neural network and geographical information system. *Renewable and Sustainable Energy Reviews*, 16(3), 1437–1449. <https://doi.org/10.1016/J.RSER.2011.11.024>
- Santana-Sarmiento, F., & Velázquez-Medina, S. (2024). A method based on GIS techniques to assess renewable energy self-consumption capacity. A case study. *Energy*, 305, 132246. <https://doi.org/10.1016/J.ENERGY.2024.132246>
- Shao, M., Han, Z., Sun, J., Xiao, C., Zhang, S., & Zhao, Y. (2020). A review of multi-criteria decision making applications for renewable energy site selection. *Renewable Energy*, 157, 377–403. <https://doi.org/10.1016/J.RENENE.2020.04.137>
- Shoei, M., Noorollahi, Y., Hajinezhad, A., & Moosavian, S. F. (2024). A review of the applications of artificial intelligence in renewable energy systems: An approach-based study. *Energy Conversion and Management*, 306, 118207. <https://doi.org/10.1016/J.ENCONMAN.2024.118207>
- Sun, L., Jiang, Y., Guo, Q., Ji, L., Xie, Y., Qiao, Q., Huang, G., & Xiao, K. (2021). A GIS-based multi-criteria decision making method for the potential assessment and suitable sites selection of PV and CSP plants. *Resources, Conservation and Recycling*, 168, 105306. <https://doi.org/10.1016/J.RESCONREC.2020.105306>
- Sun, Z., Wang, G., Li, P., Wang, H., Zhang, M., & Liang, X. (2024). An improved random forest based on the classification accuracy and correlation measurement of decision trees. *Expert Systems with Applications*, 237, 121549. <https://doi.org/10.1016/J.ESWA.2023.121549>
- Tian, J., & Ooka, R. (2024). Evaluation of solar energy potential for residential buildings in urban environments based on a parametric approach. *Sustainable Cities and Society*, 106, 105350. <https://doi.org/10.1016/J.SCS.2024.105350>
- Wang, S., Huang, X., Liu, P., Zhang, M., Biljecki, F., Hu, T., Fu, X., Liu, L., Liu, X., Wang, R., Huang, Y., Yan, J., Jiang, J., Chukwu, M., Reza Naghedi, S., Hemmati, M., Shao, Y., Jia, N., Xiao, Z., ... Bao, S. (2024). Mapping the landscape and roadmap of geospatial artificial intelligence (GeoAI) in quantitative human geography: An extensive systematic review. *International Journal of Applied Earth Observation and Geoinformation*, 128, 103734. <https://doi.org/10.1016/J.JAG.2024.103734>
- Wang, X., Xie, P., Xie, Y., & Jiang, H. (2025). Localized solar radiation zoning by combining spatially continuous estimates and Gaussian mixture models. *Journal of Atmospheric and Solar-Terrestrial Physics*, 268, 106432. <https://doi.org/10.1016/J.JASTP.2025.106432>
- Zambrano-Asanza, S., Quiros-Tortos, J., & Franco, J. F. (2021a). Optimal site selection for photovoltaic power plants using a GIS-based multi-criteria decision making and spatial overlay with electric load. *Renewable and Sustainable Energy Reviews*, 143, 110853. <https://doi.org/10.1016/J.RSER.2021.110853>
- Zambrano-Asanza, S., Quiros-Tortos, J., & Franco, J. F. (2021b). Optimal site selection for photovoltaic power plants using a GIS-based multi-criteria decision making and spatial overlay with electric load. *Renewable and Sustainable Energy Reviews*, 143, 110853. <https://doi.org/10.1016/J.RSER.2021.110853>