

1338 ANALISIS PREDIKSI RADIASI  
MATAHARI DENGAN ALGORITMA  
MACHINE LEARNING DAN  
IMPLEMENTASI BAYESIAN  
OPTIMIZATION DI PROVINSI DKI  
JAKARTA

*By Annisa Nur Oktaviani*

# ANALISIS PREDIKSI RADIASI MATAHARI DENGAN ALGORITMA MACHINE LEARNING DAN IMPLEMENTASI BAYESIAN OPTIMIZATION DI PROVINSI DKI JAKARTA

Annisa Nur Oktaviani<sup>1</sup>,

## Abstract

An increase in population leads to an increase in energy demand. Until now, energy-related problems are limited resources. Alternative renewable energy can be optimally utilized in the future. One renewable energy source is solar energy because its amount exceeds current and future energy needs. This is in line with target 7.2 in the Sustainable Development Goals (SDGs) 2030, which is to significantly increase the portion of renewable energy in the global energy mix. Indonesia has the potential for solar energy through solar radiation. However, the utilization of solar energy potential as a power plant in DKI Jakarta Province is not optimal. This study aims to predict the value of solar radiation through the daily Global Horizontal Irradiance (GHI) in DKI Jakarta using Support Vector Regression (SVR) with Bayesian Optimization and compare it with XGBoost to find the best model from the prediction results. The BO-SVR method is proven to provide good prediction results and robust on the data used because MAPE and RMSE for testing data are 0.182 and 34.412 respectively. The implementation of Bayesian Optimization in determining optimal hyperparameters in forming a prediction model has been proven to improve model performance. This research produces solar radiation predictions that provide information for the government, especially PT Perusahaan Listrik Negara (PLN) and researchers related to the characteristics of solar radiation.

**Keywords :** *Daily solar radiation, solar radiation prediction, SVR, XGboost, multivariate time series*

## Abstrak

Peningkatan populasi menyebabkan peningkatan permintaan energi. Hingga saat ini, masalah terkait energi adalah sumber daya yang terbatas. Energi alternatif terbarukan dapat dimanfaatkan secara optimal di masa depan. Salah satu sumber energi terbarukan adalah energi matahari karena jumlahnya melebihi kebutuhan energi saat ini dan masa depan. Hal ini sejalan dengan target 7.2 dalam Sustainable Development Goals (SDGs) 2030, yaitu meningkatkan porsi energi terbarukan secara signifikan dalam bauran energi global. Indonesia memiliki potensi energi matahari melalui radiasi matahari. Namun, pemanfaatan potensi energi surya sebagai pembangkit listrik di Provinsi DKI Jakarta belum optimal. Penelitian ini bertujuan untuk memprediksi nilai radiasi matahari melalui Global Horizontal Irradiance (GHI) harian di DKI Jakarta menggunakan Support Vector Regression (SVR) dengan Bayesian Optimization dan membandingkannya dengan XGBoost untuk menemukan model terbaik dari hasil prediksi. Metode BO-SVR terbukti memberikan hasil prediksi yang baik dan kuat pada data yang digunakan karena MAPE dan RMSE untuk data pengujian masing-masing adalah 0,182 dan 34,412. Penerapan Bayesian Optimization dalam menentukan hiperparameter optimal dalam membentuk model prediksi telah terbukti meningkatkan kinerja model. Penelitian ini menghasilkan prediksi radiasi matahari yang memberikan informasi bagi pemerintah, khususnya PT Perusahaan Listrik Negara (PLN) dan peneliti terkait karakteristik radiasi matahari.

**Kata kunci :** *Radiasi matahari harian, prediksi radiasi matahari, SVR, XGboost, multivariate time series*

## 1. PENDAHULUAN

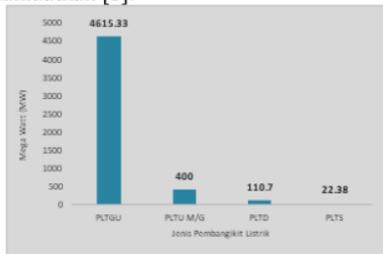
Seiring dengan terus meningkatnya pertumbuhan populasi di Indonesia, kebutuhan terhadap energi pun semakin meningkat. Energi adalah sumber daya yang diperlukan untuk berbagai macam kegiatan, seperti untuk bahan bakar, listrik, energi mekanik, dan panas [1]. Berdasarkan laporan tahunan Satuan Kerja Khusus Minda dan Gas Bumi tahun 2022, pertumbuhan produksi minyak bumi dan gas alam di Indonesia terus menurun akibat belum

ditemukannya cadangan migas lain [2]. Selain itu, energi fosil akan menyebabkan polusi udara dengan meningkatnya gas karbon yang berbahaya bagi kesehatan [3]. Berdasarkan permasalahan tersebut, sumber energi alternatif terbukti sangat dibutuhkan.

Energi surya merupakan sumber energi alternatif terbarukan yang menjanjikan dan ramah lingkungan [4]. Namun, energi surya tidak tersedia di malam hari, bergantung pada cuaca, dan memerlukan investasi awal yang tinggi [5].

Energi surya memiliki bentuk solar *thermal* untuk aplikasi pemanas dan solar *photovoltaic* untuk pembangkit listrik [3]. Energi surya dapat memenuhi kebutuhan energi global yang sejalan dengan Sustainable Development Goals 2030. Target 7.2 SDGs bertujuan untuk meningkatkan porsi energi terbarukan dalam bauran energi global sementara target 7.a berfokus pada pengembangan energi internasional, peningkatan akses terhadap penelitian dan teknologi energi hijau, serta mendorong investasi infrastruktur dan teknologi energi berkelanjutan.

Radiasi matahari adalah wujud dari potensi energi matahari sebagai alternatif sumber energi konvensional [6]. Energi ini berperan dalam menjaga keseimbangan energi di proses fisik, kimia, dan biologi [7]. Radiasi matahari dapat diubah menjadi panas atau energi listrik yang praktis untuk jangka panjang [5]. Setiap tahun, sekitar 3.400.000 Exajoule radiasi matahari mencapai permukaan bumi sedangkan konsumsi manusia hanya mencapai 450 Exajoule [4]. Artinya, radiasi matahari selama 1 jam 10 menit sudah dapat memenuhi satu tahun kebutuhan energi dalam skala global. Karena terletak di garis khatulistiwa dan beriklim tropis, terdapat banyak wilayah dengan potensi energi surya yang besar di Indonesia. Potensi energi surya berdasarkan intensitas radiasi di Indonesia berkisar 3,75 kWh/m<sup>2</sup>/hari [8]. Pada tahun 2022, potensi energi surya di Indonesia diperkirakan mencapai 3.294,4 Giga Watt tetapi hanya 0,01% (0,3 Giga Watt) yang telah dimanfaatkan [8].



**Gambar 1.** Kapasitas Pembangkit Tenaga Listrik Nasional Per Jenis Pembangkit Listrik di Provinsi DKI Jakarta Tahun 2023

Besarnya radiasi matahari di Indonesia memberikan peluang untuk Pembangkit Listrik Tenaga Surya (PLTS). Pada tahun 2022, kapasitas PLTS nasional sebesar 283,14 MW. Pada Provinsi DKI Jakarta, PLTS menduduki peringkat ketiga sebesar 22,38 MW [9]. Gambar 1 menunjukkan bahwa pemanfaatan PLTS di Provinsi DKI Jakarta paling rendah yang menandakan bahwa potensi energi surya sebagai sumber tenaga listrik belum dimanfaatkan secara optimal. Sebagai pusat perekonomian, ketergantungan DKI Jakarta terhadap energi alternatif terbarukan, seperti

radiasi matahari memberikan dampak terhadap perekonomian maupun lingkungan [10]. Peramalan yang akurat dibutuhkan untuk perencanaan dan pengelolaan energi. Namun, pengumpulan data radiasi matahari yang dilakukan BMKG masih terbatas karena kurangnya alat ukur di setiap stasiun [11]. Selain itu, metode peramalan tradisional seringkali gagal memberikan keakuratan dalam memahami dinamika cuaca.

## 2. TINJAUAN PUSTAKA

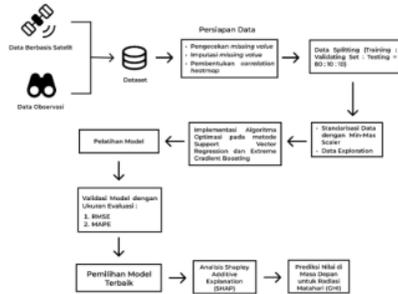
*Machine learning* adalah alat yang efektif untuk meningkatkan ketepatan dalam memprediksi kondisi lingkungan termasuk radiasi matahari karena dapat memproses data besar secara efisien melebihi metode konvensional. Algoritma ini terus berkembang meniru kecerdasan manusia melalui studi lingkungan sekitar [12]. *Support Vector Regression* (SVR) telah banyak digunakan dalam memprediksi data *time series*, termasuk radiasi matahari [15] [13]. Berdasarkan penelitian [14], SVR dengan kernel *Radial Basis Function* (RBF) memberikan hasil yang efisien dalam estimasi radiasi matahari global, menggunakan data latih selama 9 tahun dan data uji selama 4 tahun. Penelitian lain [15] menunjukkan penggunaan XGBoost dalam memperkirakan radiasi matahari melalui *Global Horizontal Irradiance* (GHI), menghasilkan generalisasi yang baik dengan nilai MAE sebesar 1.63 MJ/m<sup>2</sup>. Menurut penelitian [16], kombinasi *deep learning* dan *ridge regression* untuk memprediksi radiasi matahari setiap jam, berhasil mereduksi error sebanyak 40%.

Pemilihan variabel berperan penting dalam meningkatkan akurasi prediksi. Menurut penelitian [17], digunakan variabel seperti hari sepanjang tahun, perbedaan suhu, jam, *air quality index*, dan total tutupan awan, menghasilkan model XGBoost dengan R-squared, RMSE, dan MAPE masing-masing 0.948, 0.282, dan 0.145. Berdasarkan penelitian [18], variabel lama penyinaran matahari, suhu udara, kelembapan, serta data citra satelit digunakan untuk estimasi radiasi matahari. Selain itu, penelitian [19] menggunakan variabel temperatur untuk memprediksi radiasi matahari bulanan. Optimasi model melalui Bayesian Optimization efektif dalam meningkatkan akurasi, seperti yang ditunjukkan oleh penelitian [20] dan [21] dengan SVR dan SVM didapatkan peningkatan akurasi yang signifikan dalam prediksi. Bayesian Optimization (BO) merupakan algoritma umum untuk optimasi metode pada *machine learning*. Sehingga, penelitian ini bertujuan menggunakan SVR dengan optimasi Algoritma Bayesian dan

XGBoost untuk memprediksi nilai Global Horizontal Irradiance (GHI) harian di DKI Jakarta.

### 3. METODOLOGI PENELITIAN

#### 3.1. Tahapan Penelitian



Gambar 2. Alur Kerja Penelitian

Penelitian ini menggunakan dataset citra satelit dan data observasi dari web API, yang melalui proses *data preparation* termasuk imputasi *missing value* dan analisis korelasi visual dengan *correlation heatmap*. Data *clean* dibagi menjadi data *training* (80%), *validation* (10%), dan *testing* (10%), lalu dinormalisasi dengan *min-max scaling*. Model SVR dan XGBoost diterapkan dengan Bayesian Optimization dan dievaluasi menggunakan RMSE dan MAPE. Model terbaik dianalisis *feature importance*-nya menggunakan nilai SHAP.

#### 3.2. Pengumpulan Data

Tabel 1. Variabel Penelitian

No	Variabel Penelitian	Nama Variabel	Satuan
1	Global Horizontal Irradiance	ghi	kWh/m <sup>2</sup>
2	Air Quality Index (PM 2,5)	pm_2_5	dimensionless
3	Cloud Opacity	cloud_opacity	dimensionless
4	Zenith	zenith	°
5	Air Temperature	air_temp	°C
6	Wind Speed 10m	wind_speed_10m	(m/s)
7	Precipitable Water	precipitable_water	mm
8	Azimuth	azimuth	°
9	Relative Humidity	relative_humidity	%

10	Number of Days	Days	di antara 1 (1 January) dan 365 atau 366 (31 Desember)
11	Sunshine Duration	sun_durs	hours

Data dari variabel di atas merupakan data sekunder. Penelitian ini menggunakan data dari <https://solcast.com/> untuk variabel Cloud Opacity, Zenith, Air Temperature, Wind Speed 10m, Precipitable Water, Azimuth, Relative Humidity, dan data dari BMKG untuk variabel Air Quality Index (PM 2.5), dan Sunshine Duration di Provinsi DKI Jakarta. Data harian sebanyak 1239 observasi dari 1 Januari 2021 hingga 23 Mei 2024 berasal dari stasiun meteorologi Kemayoran dan Tanjung Priok, terdiri dari 12 atribut.

#### 3.3. Analisa Data

##### 3.3.1. K-Nearest Neighbour (KNN) Imputer

KNN imputer digunakan untuk mengimputasi nilai *missing*. KNN merupakan pengembangan dari metode NN dikarenakan sering *overfitting* [22]. Estimator kNN dengan himpunan  $k$  tetangga terdekat  $D_k = \{(X_j, Y_j, 1), j = 1, 2, \dots, k.\}$  yang diberikan  $(X, Y, 0)$ .

$$Y = \begin{cases} \arg \max_v \left\{ \sum_{(X_j, Y_j, 1) \in D_k} 1(Y_j = v) \right\} & \text{jika } Y \text{ adalah data kategorik} \\ \frac{1}{k} \sum_{j=1}^k Y_j & \text{jika } Y \text{ adalah data numerik} \end{cases} \quad (1)$$

$v$  berada dalam domain fitur target  $Y$ , dan fungsi indikator  $1(Y_j = v)$  mengembalikan nilai 1 jika argumennya benar dan 0 jika tidak.

Metode *cross-validation* dengan iterasi dapat digunakan untuk menentukan  $k$  tetangga terdekat yang optimal dalam imputasi kNN [23]. Selain itu digunakan pendekatan jarak Minkowski yang disesuaikan dengan derajat pangkatnya:

$$d(i, j) = (|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + \dots + |x_{ip} - x_{jp}|^q)^{\frac{1}{q}} \quad (2)$$

$q$  adalah koefisien Minkowski, *non-negative integer*. Jarak Minkowski dianggap jarak Euclidean saat  $q = 2$  dan sebagai jarak Manhattan ketika  $q = 1$ .

Dalam penelitian [24] yang membandingkan beberapa metode imputasi data *time series* menunjukkan hasil bahwa kNN imputer adalah

metode imputasi terbaik untuk perlakuan *missing value*.

### 3.3.2. Bayesian Optimization

Bayesian Optimization (BO) mengoptimalkan fungsi objektif yang lambat dievaluasi dengan membangun model probabilistik dan mengestimasi *hyperparameter* optimal melalui iterasi [25]. Metode ini menggunakan fungsi akuisisi untuk meningkatkan probabilitas sampel dalam pencarian [26]. Diberikan  $f(x)$  yang merupakan fungsi objektif dan  $EI(x, Q)$  merupakan peningkatan yang diharapkan berdasarkan distribusi fungsi posterior  $Q$ , maka  $EI(x, Q)$  dapat dinyatakan:

$$EI(x, Q) = E_Q \left[ \max \left( 0, \mu_Q(x_{\text{terbaik}}) - f(x) \right) \right] \quad (3)$$

$x_{\text{terbaik}}$  adalah lokasi nilai terendah rataan posterior dan  $\mu_Q(x_{\text{terbaik}})$  adalah nilai terendah rataan posterior. Penelitian menunjukkan BO meningkatkan akurasi model prediksi, seperti pada Gaussian Process Regression (hampir 99%) [27] dan BO-SVR yang mengungguli SARIMA, MLP-NN, ERT, dan Adaboost [28]. BO juga meningkatkan akurasi SVM dari 89,6% hingga 92,3% [29].

### 3.3.3. Support Vector Regression (SVR)

Support Vector Regression (SVR) berfungsi untuk mencari fungsi yang memiliki deviasi paling besar dari target aktual:

$$f(x) = \langle w, x \rangle + b, w \in X, b \in \mathbb{R} \quad (4)$$

$$w = \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i \quad (5)$$

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b \quad (6)$$

$$f(x) = w^T \varphi(x) + b \quad (7)$$

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (8)$$

$$\varphi: x \rightarrow \varphi(x) \quad (9)$$

$$K(x, y) = x^T y + c \quad (10)$$

$$K(x, y) = (\alpha x^T y + c)^d \quad (11)$$

$$K(x, y) = \exp(-\gamma \|x - y\|^2) \quad (12)$$

### 3.3.4. Extreme Gradient Boosting (XGBoost)

XGBoost terdiri dari fungsi objektif prediksi beban: kehilangan pelatihan  $L(\Theta)$  dan regularisasi  $\Omega(\Theta)$ .

$$Obj(\Theta) = L(\Theta) + \Omega(\Theta) \quad (13)$$

$\Theta$  digunakan untuk *training*, sementara  $\Omega(\Theta)$  mencegah *overfitting* dengan menghukum kompleksitas model, yang didefinisikan oleh jumlah *leaves*  $T$  (dengan konstanta  $\gamma$ ) dan penghitungan L2 standar skor daun  $\omega_j$ .

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \quad (14)$$

Fungsi objektif dari langkah ke- $t$ , yang merupakan pohon ke- $t$ :

$$\begin{aligned} Obj^{(t)} &= \sum_{i=1}^n [y_i - (\hat{y}_i^{t-1} + f_t(x_i))]^2 + \Omega(f_t) \\ &= \sum_{i=1}^n [2(\hat{y}_i^{t-1} - y_i)f_t(x_i) + f_t(x_i)^2] + \Omega(f_t) \end{aligned} \quad (15)$$

Adalah:

$$\begin{aligned} Obj^{(t)} &= \sum_{j=1}^T \left[ G_j \omega_j + \frac{1}{2} (H_j + \lambda) \omega_j^2 \right] + \gamma T \\ G_j &= \sum_{i \in I_j} g_i \text{ dan } H_j = \sum_{i \in I_j} h_i \end{aligned} \quad (16)$$

adalah turunan parsial *order* setelah mengambil ekspansi Taylor dari *loss function* pertama dan kedua. Dengan asumsi struktur pohon keputusan tetap pada XGBoost dan turunan orde pertama dari  $Obj^{(t)}$  adalah 0, nilai parameter untuk simpul daun  $\omega_j$  dapat ditemukan saat fungsi objektif mencapai titik optimal tanpa penurunan lebih lanjut di sepanjang gradien:

$$\omega_j^* = -\frac{G_j}{H_j + \lambda} \quad (17)$$

Fungsi objektifnya adalah

$$Obj^* = -\frac{1}{2} \sum_{j=1}^T \frac{G_j^2}{H_j + \lambda} + \gamma T \quad (18)$$

### 3.3.5 Korelasi Pearson

Korelasi Pearson mengukur hubungan linier antara variabel kontinu tanpa *missing value* dan jumlah observasi yang sama.

Rumus Korelasi Pearson:

$$r_{xy} = \frac{n \cdot \sum_{i=1}^n x_i \cdot y_i - (\sum_{i=1}^n x_i) \cdot (\sum_{i=1}^n y_i)}{\sqrt{[n \cdot \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2] [n \cdot \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2]}} \quad (19)$$

Nilai  $r$  berada pada interval  $[-1,1]$  dimana tanda menunjukkan arah hubungan (negatif untuk tidak searah, dan positif untuk searah), besarnya nilai  $r$  menunjukkan kekuatan hubungan 2 variabel tersebut. Dan jika bernilai 0 berarti tidak ada hubungan antar kedua variabel. Kekuatannya dapat dinilai dengan pedoman umum berikut [30].

$0.1 < |r| < 0.3$  berarti korelasi kecil atau lemah  
 $0.3 < |r| < 0.5$  berarti korelasi moderat atau sedang  
 $0.5 < |r| < 1$  maka korelasi besar atau kuat

### 3.3.6 Ukuran Evaluasi

Semakin kecil nilai MAPE dan RMSE, kesalahan peramalan akan semakin kecil, hasil peramalannya mendekati nilai aktual. Dirumuskan:

$$RMSE = \sqrt{\frac{\sum_{t=1}^n |Y_t - \hat{Y}_t|^2}{n}} \quad (20)$$

$$MAPE = \left[ \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t} \right] \times 100\% \quad (21)$$

$n$  adalah jumlah periode peramalan,  $Y_t$  adalah nilai data aktual, dan  $\hat{Y}_t$  adalah nilai peramalan. Kriteria nilai MAPE diperlihatkan pada tabel berikut [31].

**Tabel 3.** Kriteria Nilai MAPE [32]

Nilai MAPE	Kriteria
<10%	Sangat baik
10 – 20%	Baik
20 – 50%	Cukup
>50%	Buruk

### 3.3.7. Feature Importance dengan Shapley Additive Explanation (SHAP)

SHAP menghitung tingkat kepentingan fitur dengan melatih model menggunakan semua fitur

dan menghitung nilai SHAP untuk setiap fitur untuk mengidentifikasi yang paling signifikan dalam interpretasi model kompleks [33], [34].

$$\hat{\phi}_j = \frac{1}{K} \sum_k^K ((\hat{g}(x_{+j}^m) - \hat{g}(x_{-j}^m))) \quad (22)$$

$(\hat{g}(x_{+j}^m))$  adalah prediksi untuk variabel input  $x$  dengan sejumlah nilai fitur acak. Sebuah metode bernama TreeSHAP telah diajukan untuk model *gradient boosting* termasuk XGBoost [35]. Nilai estimasi TreeSHAP dapat diestimasi dengan formula berikut:

$$\Phi_{i,j} = \sum_{S \in N(i,j)} \frac{|S|! (M - |S| - 2!)}{2(M - 1)} \delta_{ij}(S) \quad (23)$$

$M$  adalah banyaknya fitur dan  $S$  adalah semua subset fitur.

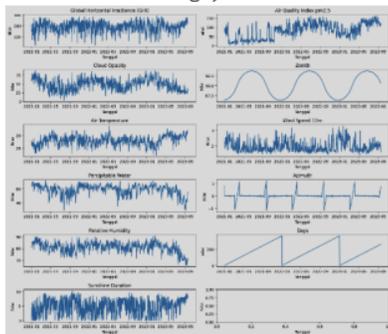
## 4. HASIL DAN PEMBAHASAN

### 4.1 Persiapan Data

Data yang diperoleh sudah berupa nilai numerik sehingga tidak perlu melakukan pengkodean ulang terhadap data yang tersedia. Sebelum membentuk model prediksi dibutuhkan proses persiapan data untuk memastikan data yang digunakan sudah siap untuk dimodelkan. Karena data diperoleh dari sumber yang berbeda sehingga perlu dilakukan penggabungan tabel dari masing – masing sumber. Dilakukan pengecekan *missing value* dari data hasil penggabungan untuk memastikan seluruh observasi dapat digunakan dalam pembentukan model. Berdasarkan pengecekan yang dilakukan diperoleh bahwa terdapat beberapa variabel yang memiliki *missing value* yaitu *air quality index* sebanyak 140 observasi, *sunshine duration* stasiun kemayoran sebanyak 14 observasi, dan *sunshine duration* Stasiun Tanjung Priok sebanyak 22 observasi. Untuk mengatasi adanya *missing value*, dilakukan imputasi dengan menggunakan *k-NN imputer*. Dengan menetapkan jumlah tetangga pada algoritma k-NN sebanyak 5 mengacu pada penelitian lain oleh [36]-[38], diperoleh tabel hasil imputasi untuk *missing value* pada data.

Setelah mengatasi masalah pada *missing value*, selanjutnya dibentuk kolom baru yang diperoleh dari rata-rata kedua variabel *sunshine duration* untuk kedua stasiun amatan. Hal ini dilakukan karena ketersediaan data terkait, untuk Provinsi DKI Jakarta hanya diamati dari

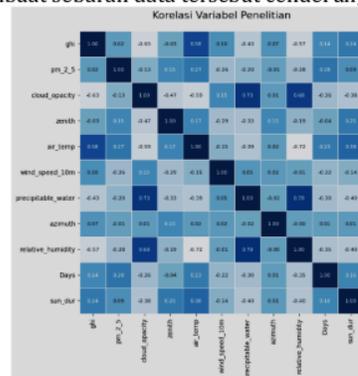
dua stasiun tersebut. Setelah membentuk variabel baru, dilakukan pendefinisian variabel yang digunakan sebagai variabel prediktor dan target pada pembentukan model. Variabel prediktor yang digunakan pada penelitian ini adalah *cloud opacity*, *air quality index pm2.5*, *zenith*, *air temperature*, *wind speed 10m*, *precipitable water*, *azimuth*, *relative humidity*, dan *sunshine duration* dengan variabel target yaitu Global Horizontal Irradiance (GHI) serta tanggal yang ditetapkan sebagai indeks masing - masing observasi. Berdasarkan variabel yang sudah didefinisikan, dilakukan *splitting* data dengan proporsi 80 : 10 : 10 yang terdiri dari *training set*, *validating set*, dan *testing set*. *Splitting* data berdasarkan tiga set data tersebut didasari atas kondisi bahwa model yang dilakukan proses optimasi sehingga membutuhkan data validasi terkait model optimal yang dibentuk. Hal tersebut membuat data *testing independent* terhadap pembentukan model sehingga akan lebih efektif untuk menguji kesesuaian model.



**Gambar 3.** Visualisasi Variabel Penelitian terhadap Periode Amatan

Berdasarkan Gambar 3 terlihat sebaran dari masing - masing variabel penelitian selama periode amatan. Variabel Global Horizontal Irradiance (GHI) yang digunakan sebagai variabel target terlihat memiliki nilai yang berfluktuasi sepanjang periode amatan. Terlihat bahwa terdapat pola musiman pada data yang membuat nilai variabel tersebut di periode tertentu memperlihatkan kondisi yang sama. Variabel *air quality index pm2.5* memperlihatkan sebaran yang berbeda untuk awal amatan. Data tersebut memberikan fluktuasi yang signifikan pada awal amatan lalu cenderung mengalami kenaikan di akhir periode amatan. Sebaran untuk variabel tersebut cenderung acak. Hal ini berbeda dengan variabel *cloud opacity* yang memperlihatkan nilai berfluktuasi namun diikuti dengan pola yang dapat dengan mudah terlihat diikuti dengan efek musiman yang teridentifikasi.

Hal ini juga diperlihatkan oleh variabel *zenith* yang memperlihatkan pola musiman yang teratur sepanjang periode amatan. Variabel *air temperature* mempunyai sebaran yang berfluktuasi serta pada periode mendekati akhir amatan nilainya cenderung mengalami kenaikan dibandingkan periode sebelumnya. Variabel *wind speed 10m* memperlihatkan fluktuasi yang cukup signifikan dengan beberapa puncak. Variabel *precipitable water* memperlihatkan trend negatif serta terdapat pola musiman pada data. Jika ditinjau kembali sebaran variabel ini cenderung sama dengan pola pada variabel *cloud opacity*. Variabel *azimuth* menunjukkan fluktuasi yang teratur sepanjang periode amatan karena menyatakan perubahan posisi matahari sepanjang periode tersebut. Variabel *relative humidity* yang cenderung berfluktuasi dan memperlihatkan penurunan pada periode akhir amatan. Variabel *days* yang merupakan nilai terurut memperlihatkan pola linier pada setiap tahun yang sama. Terakhir, variabel *sunshine duration* memperlihatkan fluktuasi yang cukup signifikan serta terdapat nilai observasi nol membuat sebaran data tersebut cenderung acak.



**Gambar 4.** Correlation Heatmap Variabel Penelitian

Setelah memperoleh sebaran data, dibentuk *correlation heatmap* terhadap variabel penelitian yang digunakan. Berdasarkan Gambar 4 terlihat bahwa keeratan linier antara variabel yang diukur menggunakan korelasi pearson cenderung rendah walaupun terdapat beberapa variabel yang memiliki keeratan linier cukup tinggi. Rendahnya korelasi pearson diduga karena keeratan data yang rendah ataupun hubungan antara variabel tidak secara linier. Selain itu terdapat variabel yang berkorelasi cukup tinggi seperti *precipitable water* dan *relative humidity* terhadap *cloud opacity*. Hal ini dapat saja terjadi karena variabel tersebut cenderung berpengaruh karena semakin tingginya kepekatan awan akan berdampak pada

peningkatan curah hujan serta peningkatan kelembaban di suatu wilayah. Selain itu jika ditinjau dari keamatan linier antara variabel *predictor* dan variabel target, variabel *cloud opacity* dan variabel *air temperature* merupakan variabel yang memiliki pengaruh terbesar terhadap radiasi matahari (GHI).

#### 4.2 Pembentukan Model

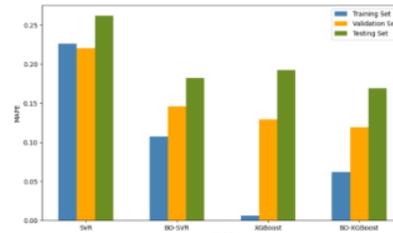
Setelah proses *data preparation* dilakukan, model dapat dibentuk menggunakan kedua metode yang ditetapkan. Pemodelan yang dilakukan melibatkan penentuan *hyperparameter* yang sesuai untuk SVR dan XGBoost menggunakan Bayesian Optimization. Untuk memperlihatkan perbedaan hasil antara model yang dioptimasi dan tanpa optimasi, pemodelan juga dilakukan langsung dengan membentuk model menggunakan metode yang ditetapkan. Hasil <sup>14</sup> luasi serta keterangan model yang dibentuk dapat dilihat pada Tabel 4.

**Tabel 4.** Evaluasi Model Machine Learning

Data	Ukuran Evaluasi	Model			
		SVR	BO-SVR	XGBoost	BO-XGBoost
Train Set	RMSE	44,339	22,402	1,665	14,227
	MAPE	0,226	0,107	0,006	0,062
Val Set	RMSE	44,339	33,02	30,29	27,179
	MAPE	0,22	0,146	0,129	0,119
Test Set	RMSE	45,46	34,412	37,54	34,051
	MAPE	0,262	0,182	0,192	0,169

Berdasarkan Tabel 4 terlihat bahwa model yang dibentuk, dievaluasi terhadap data *training*, *validating*, dan *testing* menggunakan ukuran evaluasi RMSE dan MAPE. Terlihat bahwa implementasi Bayesian Optimization (BO) dalam menentukan *hyperparameter* yang sesuai dalam pembentukan model, menghasilkan model yang lebih efisien berdasarkan ukuran evaluasi yang ditetapkan. Hal tersebut terlihat pada ukuran evaluasi model SVR yang cenderung lebih buruk dibanding BO-SVR walaupun model yang dihasilkan cukup konsisten untuk setiap bagian data karena tidak memperlihatkan simpangan yang terlalu besar pada ukuran evaluasi. Hal ini juga dapat terlihat pada ukuran evaluasi yang dihasilkan pada model XGBoost, implementasi Bayesian Optimization (BO) memberikan ukuran evaluasi yang lebih baik pada model tetapi simpangan untuk ukuran evaluasi antara data *training* terhadap *validating* dan *testing* cukup besar. Kondisi tersebut memperlihatkan kecenderungan *overfitting* yang lebih besar pada

model XGBoost dibanding model SVR. Hal ini dibuktikan dengan nilai MAPE untuk data *training* pada model XGBoost tanpa optimasi sebesar 0,008, sedangkan MAPE untuk data *testing* sebesar 0,168. Nilai tersebut cukup signifikan untuk menjelaskan perbedaan antara evaluasi model antara data yang sudah dikenali dan belum dikenali dalam prediksi yang dilakukan. Namun setelah implementasi Bayesian Optimization (BO) terlihat bahwa perbedaan antara MAPE untuk data *training* dan *testing* berkurang yang ditandai dengan jangkauan yang mengecil.



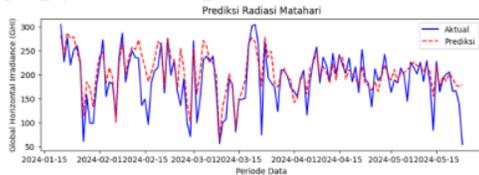
**Gambar 5.** Perbandingan Nilai MAPE Model untuk Dataset Training, Validating, dan Testing

Model yang sudah dilatih akan menghasilkan perkiraan untuk periode yang sama dengan data pengujian. Hasil yang diperkirakan akan dibandingkan dengan data pengujian untuk mengevaluasi kinerja model. Berdasarkan Gambar 5 terlihat perbedaan nilai MAPE untuk setiap dataset yang dikelompokkan berdasarkan modelnya. Pada model SVR terlihat nilai MAPE cenderung memiliki perbedaan yang tidak signifikan antara ketiga dataset. Hal tersebut memperlihatkan bahwa model yang dihasilkan cukup konsisten. Setelah diimplementasikan Bayesian Optimization (BO), MAPE dari model BO-SVR mengalami penurunan yang cukup signifikan diikuti dengan perubahan MAPE antar ketiga dataset yang tidak signifikan. Dalam pengujian model XGBoost, nilai MAPE cenderung lebih kecil dibanding kedua model sebelumnya yaitu 0,008. Namun, perubahan MAPE pada data *validating* dan *testing* cukup signifikan, hal tersebut mengindikasikan terjadinya *overfitting* pada model tersebut. Setelah mengimplementasikan Bayesian Optimization, model yang dihasilkan cenderung lebih baik saat ditinjau dari nilai MAPE dan perubahan MAPE untuk ketiga dataset. Hal tersebut membuktikan bahwa penerapan Bayesian Optimization memberikan model yang lebih baik dengan *hyperparameter* yang optimal.

**Tabel 5. Pemilihan Metode Terbaik**

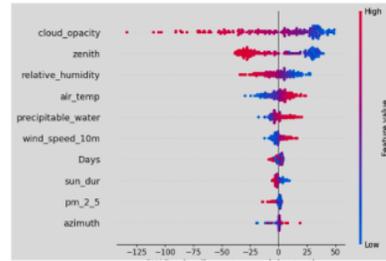
Ukuran Pemusatan	Ukuran Evaluasi	Model			
		SVR	BO-SVR	XGBoost	BO-XGBoost
Mean	RMSE	44,713	29,944	<b>23,165</b>	25,152
	MAPE	0,236	0,145	<b>0,109</b>	0,117
Standar Deviasi	RMSE	<b>0,528</b>	5,364	15,488	8,219
	MAPE	<b>0,019</b>	0,031	0,077	0,044

Dalam memilih metode yang paling baik untuk memprediksi radiasi matahari, pertimbangan standar deviasi dari RMSE dan MAPE untuk keempat model dibandingkan. Berdasarkan Tabel 5 terlihat bahwa nilai evaluasi berdasarkan rata-rata RMSE dan MAPE pada tiga dataset yang diujikan, model XGBoost memberikan performa terbaik. Sedangkan jika ditinjau dari standar deviasi untuk pengujian pada tiga dataset tersebut, model SVR memberikan hasil terbaik. Namun, model terbaik seharusnya dapat unggul pada kedua pengujian tersebut sehingga model yang dihasilkan sesuai untuk prediksi serta memberikan hasil yang konsisten. Maka, model yang dipilih sebagai model terbaik adalah BO-SVR karena memiliki rata-rata RMSE dan MAPE yang cukup rendah sehingga menjelaskan error yang dihasilkan relatif kecil serta standar deviasi dari RMSE dan MAPE yang rendah sehingga menunjukkan bahwa model tersebut memberikan hasil yang lebih konsisten karena simpangan untuk error yang terbentuk terhadap pengujian untuk ketiga dataset relatif kecil.

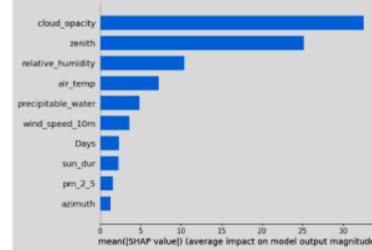


**Gambar 6. Prediksi Radiasi Matahari Menggunakan Model BO-SVR**

Berdasarkan model terbaik yang dipilih diperoleh hasil prediksi menggunakan amatan pada data *testing*. Pada Gambar 6 terlihat bahwa model prediksi yang dipilih dapat menggambarkan pola perubahan dari variabel target yaitu radiasi matahari (GHI). Namun, error dari model prediksi masih dapat diidentifikasi karena simpangan yang cukup besar dari variabel target, selain itu model yang dihasilkan cenderung *overestimate* yang dimana memperlihatkan nilai prediksi lebih tinggi dibanding nilai aktual data tersebut.

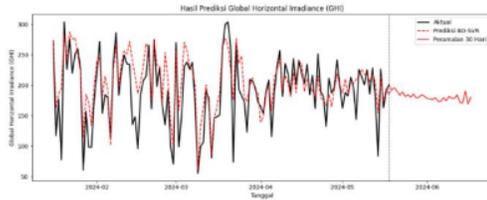


**Gambar 7. Nilai SHAP Model BO-SVR**



**Gambar 8. Rataan Nilai SHAP Model BO-SVR**

Dalam mendapatkan informasi lebih rinci dari hasil prediksi yang diperoleh, digunakan analisis Shapley Additive Explanation (SHAP) sebagai dasar interpretasi kontribusi variabel predictor terhadap variabel target dengan metode terbaik yang dipilih. Berdasarkan Gambar 8 terlihat bahwa variabel dengan rata-rata dari nilai SHAP tertinggi adalah *cloud opacity*. Nilai SHAP positif menyatakan bahwa kondisi dimana variabel target mengalami peningkatan, sehingga nilai rendah untuk variabel *cloud opacity* cenderung memberikan peningkatan pada variabel GHI. Selain itu juga dapat dilihat pada Gambar 7, dimana sebaran yang terbentuk pada plot masing-masing variabel, menggambarkan besar pengaruh nilai variabel tersebut terhadap perubahan nilai variabel target yaitu GHI. Hal tersebut juga berarti jika nilai SHAP memusat pada nilai 0, maka variabel tersebut memberi pengaruh yang relatif kecil terhadap variabel target. *Zenith* merupakan variabel selanjutnya yang memiliki rata-rata dari nilai SHAP yang cukup tinggi, dimana nilai rendah pada variabel tersebut memberikan peningkatan pada variabel GHI. Berbeda dengan variabel sebelumnya, variabel *air temperature* memberikan peningkatan pada variabel GHI menggunakan nilai tinggi pada variabel tersebut.



**Gambar 9.** Forecasting GHI Berdasarkan Model BO-SVR

Berdasarkan model serta metode terbaik yang telah dipilih, dilakukan prediksi terhadap variabel GHI untuk tiga puluh observasi selanjutnya. Prediksi yang dilakukan diperoleh dengan memodelkan masing-masing variabel independent menggunakan lag variabel terkait menggunakan metode BO-SVR. Berdasarkan masing-masing model, diprediksi sebanyak tiga puluh observasi selanjutnya sehingga akan terbentuk daftar variabel independent tanpa label. Berdasarkan data tersebut dilakukan pemodelan dengan menggunakan model terbaik sehingga diperoleh hasil prediksi variabel GHI untuk tiga puluh observasi selanjutnya. Berdasarkan Gambar 9 diperoleh hasil prediksi tiga puluh observasi selanjutnya untuk variabel GHI. Berdasarkan pola yang diperoleh sebelumnya, terlihat bahwa prediksi model BO-SVR cenderung dapat menangkap trend pada data GHI sehingga mengacu pada hasil peramalan dapat diputuskan bahwa nilai GHI akan naik pada periode yang menunjukkan kenaikan pada grafik yang disajikan. Namun, pada hasil peramalan yang diperoleh terlihat kecenderungan nilai GHI yang diprediksi tidak menunjukkan fluktuasi yang setara seperti data aktual.

**Tabel 6.** Prediksi GHI 30 Hari Selanjutnya

Tanggal	Hasil Prediksi
19 Mei 2024	192.447
20 Mei 2024	195.807
21 Mei 2024	190.082
...	...
15 Juni 2024	190.179

Berdasarkan Tabel 6, ditunjukkan nilai periode optimal hasil prediksi GHI selama tiga puluh hari kedepan. Diperoleh bahwa nilai optimal terdapat pada tanggal 19, 20, 21 Mei serta tanggal 15 Juni 2024.

## 5. KESIMPULAN DAN SARAN

Radiasi matahari memiliki potensi besar sebagai sumber energi terbarukan untuk produksi listrik di DKI Jakarta. Model prediksi

radiasi matahari yang memanfaatkan variabel predictor seperti cloud opacity, air quality index pm2.5, zenith, air temperature, wind speed 10m, precipitable water, azimuth, relative humidity, sunshine duration, dan days menunjukkan bahwa metode BO-SVR memberikan hasil prediksi yang baik dan robust. Evaluasi menggunakan MAPE dan RMSE menunjukkan nilai masing-masing sebesar 0,182 dan 34,412. Implementasi Bayesian Optimization dalam menentukan hyperparameter optimal meningkatkan performa model, sementara analisis Shapley Additive Explanation (SHAP) mengungkapkan bahwa variabel cloud opacity memiliki pengaruh terbesar, sedangkan azimuth memiliki pengaruh terendah.

Hasil penelitian ini memberikan informasi penting bagi pemerintah dan PT PLN untuk mengembangkan teknologi Pembangkit Listrik Tenaga Surya (PLTS). Pemerintah dapat memanfaatkan data ini untuk memperkirakan periode optimal untuk penggunaan energi matahari, meskipun data radiasi matahari berfluktuasi. Nilai prediksi GHI menunjukkan kecenderungan tinggi mendekati akhir bulan, yang dapat dimanfaatkan sebagai periode optimal untuk produksi listrik. Selain itu, variabel seperti cloud opacity dan zenith yang memiliki efek musiman perlu diperhatikan dalam menentukan periode optimal pemanfaatan energi matahari. Implementasi machine learning dengan Bayesian Optimization terbukti efektif, memberikan opsi bagi peneliti selanjutnya untuk memanfaatkan algoritma machine learning dalam prediksi radiasi matahari.

# 1338 ANALISIS PREDIKSI RADIASI MATAHARI DENGAN ALGORITMA MACHINE LEARNING DAN IMPLEMENTASI BAYESIAN OPTIMIZATION DI PROVINSI DKI JAKARTA

---

ORIGINALITY REPORT

---

# 5%

SIMILARITY INDEX

---

## PRIMARY SOURCES

---

1	Maram Magdy Salem. "Bayesian Analysis for Dependent Progressively Censored Weibull Competing Risks using Copulas", International Journal of Reliability, Quality and Safety Engineering, 2023 Crossref	33 words — 1%
2	<a href="http://journal.aritekin.or.id">journal.aritekin.or.id</a> Internet	28 words — 1%
3	<a href="http://jurnal.umk.ac.id">jurnal.umk.ac.id</a> Internet	11 words — < 1%
4	<a href="http://jwcn-urasipjournals.springeropen.com">jwcn-urasipjournals.springeropen.com</a> Internet	11 words — < 1%
5	<a href="http://is.muni.cz">is.muni.cz</a> Internet	10 words — < 1%
6	<a href="http://eprints.undip.ac.id">eprints.undip.ac.id</a> Internet	9 words — < 1%
7	<a href="http://jurnal.darmajaya.ac.id">jurnal.darmajaya.ac.id</a> Internet	9 words — < 1%
8	<a href="http://www.bpk.go.id">www.bpk.go.id</a> Internet	9 words — < 1%

---

9	<a href="http://www.coursehero.com">www.coursehero.com</a> Internet	9 words — < 1%
10	<a href="http://www.datatistika.com">www.datatistika.com</a> Internet	9 words — < 1%
11	<a href="http://www.lib.ui.ac.id">www.lib.ui.ac.id</a> Internet	9 words — < 1%
12	<a href="http://zh.scribd.com">zh.scribd.com</a> Internet	9 words — < 1%
13	<a href="http://docplayer.info">docplayer.info</a> Internet	8 words — < 1%
14	<a href="http://id.123dok.com">id.123dok.com</a> Internet	8 words — < 1%
15	<a href="http://jurnal.ugm.ac.id">jurnal.ugm.ac.id</a> Internet	8 words — < 1%
16	<a href="http://nanopdf.com">nanopdf.com</a> Internet	8 words — < 1%
17	<a href="http://repositori.uin-alauddin.ac.id">repositori.uin-alauddin.ac.id</a> Internet	7 words — < 1%
18	<a href="http://repository.unair.ac.id">repository.unair.ac.id</a> Internet	7 words — < 1%
19	Josef Spiegel. "Sexual Abuse of Males - The SAM Model of Theory and Practice", Routledge, 2019 Publications	6 words — < 1%
20	<a href="http://moam.info">moam.info</a> Internet	6 words — < 1%

---

---

EXCLUDE QUOTES      OFF

EXCLUDE BIBLIOGRAPHY   OFF

EXCLUDE SOURCES      OFF

EXCLUDE MATCHES      OFF