

Random Forest Regression Untuk Prediksi Produksi Daya Pembangkit Listrik Tenaga Surya

Ardianto⁽¹⁾, Agus Budi Raharjo⁽²⁾, Diana Purwitasari⁽³⁾

¹PLN Unit Pelaksana Pengatur Distribusi Makassar
Jl. Serui No.5A, Makassar, Sulawesi Selatan, Indonesia

²Teknik Informatika, Institut Teknologi Sepuluh Nopember

³Magister Manajemen Teknologi, Institut Teknologi Sepuluh Nopember
Jl. Teknik Kimia, Keputih, Kec. Sukolilo, Kota SBY, Jawa Timur, Indonesia

Email: ¹ardianto.aja@gmail.com, ²agus.budi@its.ac.id, ³diana@if.its.ac.id

Tersedia Online di

<http://www.jurnal.unublitar.ac.id/index.php/briliant>

Sejarah Artikel

Diterima pada 1 Juli 2022
Disetujui pada 25 Juli 2022
Dipublikasikan pada 22 November 2022
Hal. 1058-1075

Kata Kunci:

PLTS; Peramalan; Jangka pendek; Pembelajaran Mesin; *Random Forest*

DOI:

<http://dx.doi.org/10.28926/briliant.v7i4.1036>

Abstrak: Pembangkit listrik tenaga surya (PLTS) menjadi solusi yang paling populer dan diterapkan di banyak negara. Namun interkoneksi PLTS ke sistem jaringan transmisi listrik menghadirkan permasalahan kepada operator jaringan dikarenakan memiliki sifat fluktuasi dalam menghasilkan energi listrik. Faktor-faktor yang berpotensi mempengaruhi sifat fluktuasi energi listrik adalah meteorologi dan parameter cuaca. Salah satu langkah mitigasi untuk mengatasi kondisi tersebut yaitu dengan memprediksi produksi keluaran daya PLTS. Penelitian ini mengajukan metode prediksi produksi daya dengan pra-proses data, penerapan model regresi, serta penentuan skenario uji coba. Data historis PLTS berasal dari sistem SCADA selama setahun yang terdiri atas faktor nilai produksi keluaran daya, radiasi, suhu lingkungan, suhu peralatan, dan kecepatan angin. Data yang telah diolah selanjutnya dimodelkan menggunakan algoritma *Random Forest Regression* (RFR). Dalam proses pemodelan dilakukan skenario pengaturan beberapa parameter, seperti proses perbaikan hilang rekam, normalisasi data dan filter produksi. Evaluasi dilakukan dengan menganalisis perbandingan kinerja setiap algoritma beserta kombinasi skenarionya. Hasil eksperimen menunjukkan bahwa RFR mempunyai kinerja tinggi dengan nilai R^2 sebesar 0.9679 dan RMSE sebesar 0.0438. Pemilihan skenario yang tepat terbukti memberi peningkatan kinerja akurasi sebesar RFR 2,90%.

PENDAHULUAN

Perkembangan teknologi ketenagalistrikan semakin berkembang, terutama tantangan untuk mengubah energi fosil menjadi sumber energi baru terbarukan, hal ini menjadi perhatian utama sebagian besar negara karena alasan lingkungan, dan menjaga kelestarian alam (Cebrucean, et al., 2020). Pemerintah Indonesia telah mengumumkan akan mencapai dekarbonisasi pada tahun 2060, komitmen tersebut dapat dilihat dalam Rencana Usaha Penyediaan Tenaga Listrik Perusahaan Listrik Negara (RUPTL) PLN 2021-2030 dengan porsi kapasitas pembangkit EBT sebesar 51,6% atau 20,923 MW pada 2030 (Menteri ESDM & PLN, 2021). Rencana ini memerlukan kesinambungan antara penambahan kapasitas energi terbarukan dan

dekomisioning pembangkit berbahan bakar fosil. Energi surya menjadi sumber yang paling menjanjikan sebagai penghasil listrik untuk perumahan, komersial, dan industri (Wan, et al., 2015). Dalam praktiknya, variasi variabel meteorologis yang berubah seiring ruang dan waktu membuat produksi listrik dari sistem fotovoltaik menjadi tidak pasti atau berfluktuasi. Variasi produksi listrik yang fluktuatif sangat dipengaruhi oleh kondisi atmosfer, radiasi matahari, kondisi cuaca, suhu, pergerakan awan, kecepatan angin, karakteristik dari panel surya yang digunakan, dan faktor eksternal lainnya. Pembangkit listrik tenaga surya (PLTS) skala besar yang terhubung ke sistem jaringan menimbulkan masalah jaringan yang signifikan seperti stabilitas sistem, keandalan, keseimbangan kapasitas, daya reaktif, dan respons frekuensi. Diskontinuitas ini menghadirkan tantangan besar bagi operator dalam memprediksi parameter jaringan pada tahap perencanaan atau pengoperasian sistem tenaga listrik secara *real time* (Sobrina, et al., 2018).

Dalam operasi sistem tenaga, operator jaringan transmisi harus merencanakan operasi dengan menetapkan jadwal *unit commitment* dan keputusan *economic dispatch* untuk hari berikutnya (Csereklyei, et al., 2019). Karena energi PV sangat fluktuatif dan dapat bervariasi dari nol hingga ratusan persen, maka operator jaringan harus melakukan mitigasi untuk mengatasi skenario apa pun saat mentransmisikan energi dari pembangkit listrik tenaga surya (Mohy-ud-din, et al., Juni 2019). Oleh karena itu, penting bagi operator jaringan untuk memperkirakan kapasitas operasi pembangkit listrik tenaga surya yang terhubung satu sama lain dalam jaringan, agar dapat secara optimal menemukan keseimbangan antara produksi dan permintaan beban di seluruh jaringan, inersia dan cadangan putar pembangkit, menjaga keandalan, dan efisiensi manajemen jaringan. Sehingga profitabilitas operasi jaringan juga bergantung pada kemampuan operator untuk memprediksi fluktuasi produksi PLTS secara akurat.

Banyak teknik peramalan PV yang telah dikembangkan, Sobrina dkk (Sobrina, et al., 2018) dalam penelitiannya, mereka menganalisis dan mengelompokkan beberapa teknik peramalan PLTS. Teknik peramalan tersebut dikelompokkan menjadi tiga metode utama, yaitu metode deret waktu statistik, metode fisik dan metode sintetik (gabungan). Metode statistik meliputi lima sub-model yaitu, (i) Artificial Neural Network (ANN), (ii) Support Vector Machine (SVM), (iii) Markov Chain, (iv) Autoregressive, and (v) model regresi. Metode statistik ini terutama didasarkan pada data historis, kemudahan implementasi dan kemampuan untuk mengekstrak data masa lalu untuk peramalan deret waktu. Metode fisik terdiri dari tiga sub-model yaitu, (i) Numerical Weather Prediction (NWP), (ii) Sky Imagery, dan (iii) model Satellite-Imaging. Model fisik ini didasarkan pada interaksi antara keadaan fisik dan gerakan dinamis radiasi matahari yang terjadi di atmosfer. Metode gabungan (*ensemble methode*) merupakan gabungan dari metode statistik dan metode fisik. Konsepnya adalah untuk menggabungkan model yang berbeda dengan fitur unik untuk mengatasi keterbatasan model individu sehingga meningkatkan kinerja peramalan (Leva, et al., 2017). Teknik ini juga mampu menggabungkan pendekatan linier dan non-linier.

Peramalan PV berbasis *Machine Learning* oleh Abinet dkk (Abinet, et al., 2018) melakukan peramalan jangka pendek produksi PLTS di Beijing untuk satu hari kedepan, menggunakan dataset timeseries histori produksi dari sistem SCADA dan data NWP meteorologi selama setahun dengan resolusi 10 menit. Model yang digunakan adalah hybrid gabungan dari SVM, PSO dan WT. Peramalan dianalisa

melalui pengelompokan jenis musim (dingin, semi, panas, hujan), beberapa kerangka waktu (3,6,12,24 jam) dan kinerja dibandingkan dengan beberapa model melalui nilai MAPE, SSE, RMSE, SDE. Kinerja model yang diusulkan berhasil dan efektif dengan MAPE harian dan NMAE masing-masing memiliki nilai rata-rata 4,22% dan 0,4%, mengungguli tujuh strategi prediksi lainnya sementara waktu komputasi rata-rata lebih kecil dari 15 detik. Dengan demikian, hasil numerik yang ditunjukkan memverifikasi efektivitas strategi yang diusulkan untuk prediksi tenaga surya PV jangka pendek.

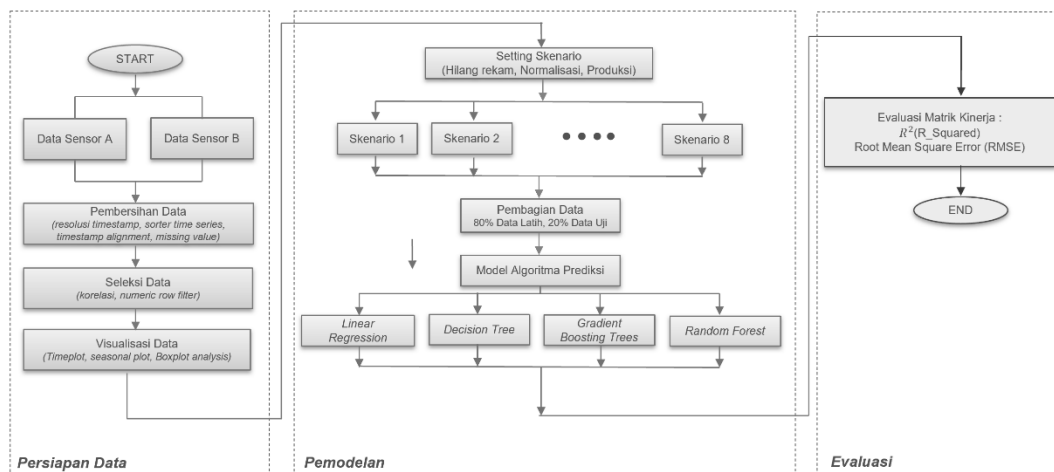
Di sisi lain, beberapa penelitian telah menggunakan metode gabungan, seperti yang dilakukan oleh Tesfaye (Abinet, et al., 2018), yang mengusulkan kombinasi transformasi wavelet, particle swarm optimization dan support vector machine (Hybrid WT-PSO-SVM) untuk peramalan daya PLTS jangka pendek (satu hari kedepan) dalam sistem grid. Data untuk model yang digunakan adalah rekam data nilai daya pembangkit dari pembacaan SCADA (*Supervisory Control and Data Acquisition*) dikombinasikan dengan data meteorologi NWP. Kinerja dibandingkan dengan 7 model algoritma lainnya dan menghasilkan peningkatan akurasi peramalan.

Dalam penelitian ini dilakukan peramalan berbasis pembelajaran mesin untuk satu periode dengan dataset berupa data *histori* dari pembacaan sensor melalui sistem SCADA di PLTS pada jaringan grid sub-sistem Sulutgo (Sulawesi bagian utara – Gorontalo) selama setahun dari Januari s.d. Desember 2021. Penelitian dilakukan untuk memprediksi produksi PLTS menggunakan algoritma Random Forest Regression (RFR) dan dibandingkan dengan algoritma *Linear Regression* (LR), *Decision Tree Regression* (DTR) dan *Gradient Boosting Trees* (GTR). Pemilihan keempat metode pembandingan tersebut mewakili model regresi dalam *supervised machine learning* untuk data deret waktu. Dalam proses pemodelan juga dilakukan skenario pengaturan beberapa parameter yaitu proses perbaikan hilang rekam (data lompat), proses normalisasi data dan proses filter produksi. Evaluasi kinerja peramalan dilakukan dengan menganalisis perbandingan setiap algoritma beserta kombinasi skenario pengaturan parameter melalui indikator akurasi berupa R-Squared (R^2) dan RMSE.

Berdasarkan latar belakang permasalahan di atas, rumusan masalah pada penelitian ini adalah; bagaimana persiapan data dan visualisasi data SCADA untuk peramalan jangka pendek PLTS, bagaimana pembuatan modelnya, dan bagaimana pengukuran kinerja hasil peramalannya. Tujuan yang ingin dicapai adalah menemukan model peramalan produksi PLTS dari empat metode dan skenario berdasarkan nilai akurasi tertinggi. Diharapkan dengan implementasi model terbaik tersebut, operator jaringan dapat melakukan pengaturan komposisi jaringan dengan lebih presisi terkait kontribusi penjadwalan *unit commitment* pembangkit serta *economic dispatch*, yang menjadikan biaya pokok penyediaan (BPP) energi listrik akan lebih efisien. Sehingga operator mendapatkan efisiensi bisnis berupa manfaat finansial dari penyediaan energi listrik.

METODE

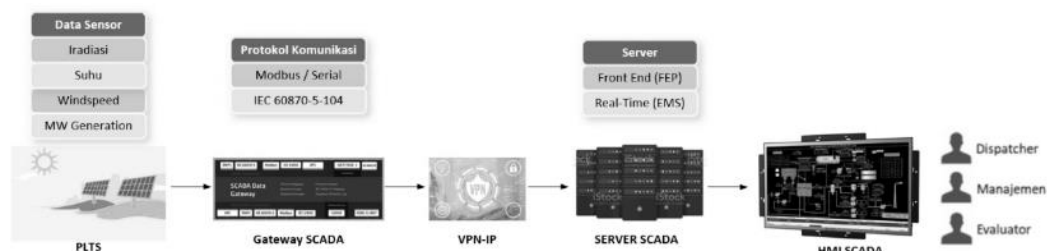
Pada bab ini akan menjelaskan metode proses tahapan untuk membuat model peramalan prediksi daya pembangkit. Tahapan ini terdiri dari 3 subbab yaitu persiapan data, pemodelan, dan evaluasi seperti yang terlihat pada gambar 1.



Gambar 1. Diagram alir proses pemodelan PLTS

1) Persiapan Data

Tahap awal dari persiapan data adalah melakukan pengumpulan data. Dataset penelitian ini adalah gabungan pembacaan dari masing-masing jenis sensor di dua titik (lokasi A dan lokasi B) yang berada dalam satu lokasi PLTS, hasil pembacaan sensor tersebut kemudian akan dikumpulkan pada suatu perangkat gateway yang kemudian akan dikirimkan melalui jaringan VPN-IP dengan menggunakan standar protokol komunikasi sistem SCADA. Sesuai ilustrasi pada gambar 2, server SCADA akan mengolah sekaligus memvisualisasikan data sensor pada HMI (*Human Machine Interface*) yang digunakan untuk pengaturan sistem, monitoring dan evaluasi oleh operator (dispatcher), manajemen dan evaluator. Data pembacaan sensor akan disimpan secara terus menerus pada server khusus database yang disebut *historical server*.



Gambar 2. Proses pengumpulan datas pada sistem SCADA

Dataset diambil dari PLTS di jaringan sub-sistem Sulutgo (Sulawesi bagian utara dan Gorontalo) selama setahun sejak Januari s.d. Desember 2021, sebanyak 524.287 baris dan 10 kolom. Kumpulan data dibagi menjadi fitur dan target, sebagai fitur adalah pembacaan dari sensor cuaca seperti sensor radiasi, suhu lingkungan, kecepatan angin, dan suhu modul panel surya. Sedangkan yang menjadi target adalah data riwayat produksi keluaran daya (*Generation*). Struktur dataset untuk penelitian ini ditunjukkan pada Tabel 1.

Tabel 1. Struktur dataset dan variable PLTS

No	Nama Field	Unit	Type	Keterangan	Variabel
1	Timestamp		Datetime	Waktu perekaman data	Fitur
2	Generation	MW	Numerik	Nilai produksi pembangkit	Target
3	A - Radiation	W/m2	Numerik	Radiasi lokasi A	Fitur
4	A - Ambient Temp	C	Numerik	Suhu sekitar lokasi A	Fitur
5	A - Wind speed	m/s	Numerik	Kecepatan angin lokasi A	Fitur
6	A - Module Temp	C	Numerik	Suhu modul lokasi A	Fitur
7	B - Radiation	W/m2	Numerik	Radiasi lokasi B	Fitur
8	B - Ambient Temp	C	Numerik	Suhu sekitar lokasi B	Fitur
9	B - Wind speed	m/s	Numerik	Kecepatan angin lokasi B	Fitur
10	B - Module Temp	C	Numerik	Suhu modul lokasi B	Fitur

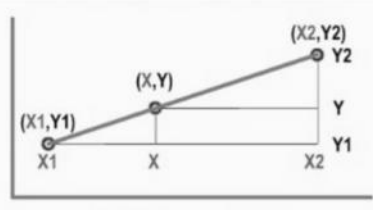
Selanjutnya dilakukan pengolahan data yang terdiri dari proses pembersihan, pemilihan, dan visualisasi data. Proses pembersihan meliputi manipulasi penyeragaman resolusi waktu data pembacaan sensor menjadi permenit dan pengurutan berdasarkan waktu pembacaan. Hasil observasi data mentah menunjukkan bahwa resolusi waktu (timestamp) dari pembacaan hasil rekaman masing-masing sensor bervariasi dari 5 detik s.d. 1 menit. Sehingga perlu dilakukan penyeragaman dengan memecah format timestamp menjadi tanggal, jam, menit dan detik. Selanjutnya dilakukan pengelompokan pada setiap tanggal-jam-menit dengan mengambil nilai rata-rata (metode *mean*) dari nilai pembacaan pada semua sensor pada menit yang sama, dengan rumus yang ditunjukkan pada Persamaan (1).

$$\bar{x} = \frac{\sum x}{n} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (1)$$

Setelah itu dilakukan pengurutan data sesuai urutan waktunya dikarenakan dataset adalah data deret waktu sehingga penting untuk melakukannya agar didapat informasi aktual sesuai situasi di lapangan. Dataset yang telah diurutkan kemudian dilakukan penyelerasan timestamp untuk memastikan bahwa SCADA ataupun sensor selalu melakukan pembacaan terus menerus setiap menitnya untuk memastikan tidak terjadi hilang rekam (data lompat) yang dikarenakan oleh gangguan sensor ataupun gangguan sistem lainnya. Validasi dilakukan dengan mengurutkan timestamp per menit. Jika menit hilang, baris baru dibuat dengan nilai stempel waktu sesuai dengan menit yang hilang sampai dengan urutan menit yang tidak hilang, dengan isian nilai variabel dikosongkan atau dianggap nilai hilang (*missing value*), seperti yang ditunjukkan pada gambar 3a. Pengisian nilai yang kosong dipilih menggunakan metode *linear interpolation*, karena sifat dari pembacaan sensor yang selalu naik atau turun secara linear, seperti terlihat pada gambar 3b. Sehingga nilai yang hilang dihitung dari nilai baris sebelum dan sesudah yang tidak hilang berdasarkan nilai antara dua persamaan linier atau disebut juga persamaan garis lurus, seperti yang ditunjukkan pada Persamaan (2).

NEWTIMESTAMP	D	Gene...	D	...	D	A...
2021-03-15T17:44	0.166	22	29.08			
2021-03-15T17:45	0.157	21	29.01			
2021-03-15T17:47	0.133	18	28.87			
2021-03-15T17:48	0.122	17	28.81			
2021-03-15T17:49	0.109	15	28.75			
2021-03-15T17:50	0.097	14	28.71			
2021-03-15T17:51	0.086	12	28.69			
2021-03-15T17:52	0.075	11	28.67			
2021-03-15T17:53	0.067	10	28.63			
2021-03-15T17:54	0.061	9	28.61			
2021-03-15T17:55	0.055	8	28.57			
2021-03-15T17:56	0.049	7	28.55			
2021-03-15T17:57	0.043	7	28.51			
2021-03-15T17:58	0.035	6	28.48			
2021-03-15T18:00	0.02	4	28.42			
2021-03-15T18:02	0.005	3	28.35			
2021-03-15T18:03	0	2	28.32			
2021-03-15T18:04	0	1	28.28			

(a) Penyalarsan waktu rekam sensor



(b) Persamaan garis lurus (linear interpolation)

Gambar 3. Penanganan data hilang rekam (data lompat)

$$\frac{(X - X1)}{X2 - X1} = \frac{(Y - Y1)}{(Y2 - Y1)} \quad Y = Y1 + \frac{(Y - Y1)}{(Y2 - Y1)}(Y2 - Y1) \quad (2)$$

Tahap selanjutnya adalah seleksi data, tahap ini merupakan proses menyusun dataset yang disesuaikan dengan kebutuhan penelitian. Pertama dilakukan cek korelasi antara setiap fitur terhadap target. Pengecekan korelasi menggunakan Pearson yang merupakan metode statistik guna melihat kedekatan hubungan secara linier antara dua variabel (x,y). Koefisien korelasi dapat bernilai positif (+) dan negatif (-), dimana nilai tersebut menunjukkan arah hubungan. Koefisien korelasi bertanda positif memiliki arti adanya hubungan berbanding lurus, bertanda negatif memiliki arti hubungan berbanding terbalik dan bernilai nol berarti tidak memiliki hubungan (Yeager, 2022). Proses terakhir adalah seleksi data adalah melakukan visualisasi dataset untuk memudahkan dalam melakukan analisis deskriptif terhadap dataset yang akan dimodelkan. Sehingga bisa ditentukan data mana saja yang akan digunakan dalam pemodelan, visualisasi yang digunakan adalah statistik, grafik dan analisis boxplot.

2) Pemodelan

Tahap awal dalam pemodelan adalah pengaturan parameter untuk menentukan skenario pemodelan dan pengujian. Ada tiga parameter yang diamati dalam penelitian ini, yaitu:

- Ada tidaknya proses perbaikan data hilang rekam (data lompat), proses ini memperbaiki lompat data dengan mengisi baris baru pada timestamp yang hilang.
- Ada tidaknya proses filter data hanya saat PLTS memproduksi (*generation filter*), proses ini menghilangkan semua baris data untuk semua nilai target yang kurang dari nol, yaitu saat tidak mengeluarkan keluaran daya. Selain karena periode pemeliharaan, juga terdapat nilai negatif yang disebabkan oleh konfigurasi PLTS yang memungkinkan dilakukan impor daya listrik dari eksternal, terutama saat tidak memproduksi dan untuk kepentingan catu daya peralatan. Sehingga data tersebut perlu dipertimbangkan untuk dikeluarkan agar bisa diketahui pengaruhnya terhadap akurasi hasil model prediksi.

- c) Ada tidaknya proses normalisasi dataset, proses ini membuat semua variabel (fitur dan target) memiliki rentang nilai yang sama. Dalam penelitian ini digunakan metode normalisasi Min-Max, menggunakan rumus pada persamaan (3), metode ini menghasilkan nilai baru hasil normalisasi antara 0 s.d. 1.

$$X_{new} = \frac{X_{old} - X_{min}}{X_{max} - X_{min}} \quad (3)$$

Pengaturan skenario ini bertujuan untuk menganalisis dan mengevaluasi masing-masing pengaruh tiap parameter terhadap akurasi yang dihasilkan oleh model prediksi. Dari kombinasi pengaturan ketiga parameter tersebut didapatkan delapan skenario dengan masing-masing pengaturan parameternya bisa dilihat pada Tabel 2.

Tabel 2. Variasi skenario pemodelan

Skenario (S)	Perbaikan Hilang Rekam	Normalisasi Data	Filter Produksi
S1	Ada	Ada	Ada
S2	Ada	Ada	Tidak Ada
S3	Ada	Tidak Ada	Tidak Ada
S4	Tidak Ada	Tidak Ada	Tidak Ada
S5	Tidak Ada	Ada	Ada
S6	Tidak Ada	Tidak Ada	Ada
S7	Ada	Tidak Ada	Ada
S8	Tidak Ada	Ada	Tidak Ada

Dataset yang telah diatur sesuai pilihan skenario, kemudian dilakukan pembagian data menjadi 80% data latih dan 20% data uji. Selanjutnya dieksekusi program menggunakan algoritma RFR, GBT, DTR dan LR. RFR merupakan metode gabungan yang terdiri dari lebih dari satu *decision tree* yang saling independen. Pemodelan menggunakan RFR dinilai lebih memberikan kinerja yang lebih baik jika dibandingkan hanya dengan menggunakan satu *decision tree* (Müller & Guido, 2016). Hasil prediksi yang dibuat oleh RFR adalah hasil terbaik dari setiap DTR, diambil rata-rata hasil untuk kasus prediksi regresi dan voting terbanyak untuk kasus klasifikasi. Model RFR yang terdiri dari N trees dirumuskan pada Persamaan (4).

$$l(y) = \operatorname{argmax}_c \left(\sum_{n=1}^N I_{h_n(y)=c} \right) \quad (4)$$

Dimana I adalah fungsi indikator dan h_n adalah *tree* ke-n. RFR memiliki mekanisme internal yang menyediakan estimasi kesalahan yang disebut estimasi kesalahan *out-of-bag* (OOB). Pengambilan sampel *bootstrap* hanya menggunakan 2/3 dari data asli. Sepertiga sisanya dikategorikan berdasarkan pohon yang dibentuk dan digunakan untuk menguji kinerjanya. Estimasi kesalahan OOB adalah rata-rata kesalahan prediksi untuk setiap kasus pelatihan y menggunakan pohon yang tidak menyertakan y dalam sampel *bootstrap* (Dimitris, et al., 2014).

Sebagai salah satu pembanding yaitu algoritma GBT, merupakan algoritma pembelajaran mesin yang akurat untuk memprediksi variable target dengan menggabungkan perkiraan satu set model yang lebih sederhana dan lebih

lemah sehingga nantinya akan terbentuk prediksi akhir yang lebih akurat (Chen & Guestrin, 2016). Algoritma ini merupakan gabungan dari dua algoritma yaitu algoritma pohon regresi dan *boosting* yang menggunakan peningkatan akurasi prediktor. Setiap pohon regresi akan memetakan titik data input kepada salah satu daun yang diinisiasikan pada skor tertentu dan pelatihan tersebut dilakukan secara berulang, gradien yang dimaksud adalah dengan penambahan pohon baru yang digunakan untuk mendapatkan skor tertentu dengan tetap memprediksi residu/kesalahan pada pohon sebelumnya kemudian dikombinasikan sehingga dapat membuat prediksi akhir. Pada algoritma RFR dan GBT mempunyai berbagai *hyperparameter* yang membutuhkan penyetelan untuk kinerja algoritma yang optimal. Optimasi *hyperparameter* adalah ilmu menyetel *hyperparameter* dari algoritma untuk mendapatkan kinerja yang optimal dimana algoritma yang berbeda memiliki tipe *hyperparameter* yang berbeda pula (Budholiya, et al., 2020). Beberapa parameter utama *hyperparameter* untuk meningkatkan algoritma RFR dan GBT yaitu:

a) *n_estimator*

adalah jumlah iterasi dalam pelatihan. Dalam penelitian ini RFR dan GBT memakai *n_estimator* berjumlah 100.

b) *tree_depth*

adalah kedalaman maksimum pohon pembelajaran. Dalam penelitian ini untuk RFR tidak diatur nilainya, GBT diatur sebanyak 4.

c) *Minimum_node_size*

Jumlah minimum node anak dalam setiap node decision. Dalam penelitian ini tidak diatur nilainya.

d) *learning_rate*

adalah parameter yang sangat penting yang perlu di sesuaikan, karena sangat mempengaruhi kinerja model. Dalam penelitian ini, *learning_rate* GBT nilainya diatur 0,1.

3) Evaluasi

Analisa dan evaluasi hasil prediksi dari model dilakukan dengan menganalisis perbandingan kinerja algoritma RFR dengan algoritma LR, DTR, dan GBT beserta kombinasi skenario pengaturan parameter melalui indikator akurasi berupa R-Squared (R^2) dan RMSE. Pemilihan ketiga algoritma pembanding tersebut mewakili model regresi dalam *supervised machine learning* untuk data deret waktu, dan juga sudah banyak referensi menggunakannya dalam peramalan daya listrik (Sobrina, et al., 2018). Akurasi setiap algoritma beserta skenarionya kemudian dikelompokan berdasarkan efisiensi peramalannya dan diberikan pembobotan untuk dianalisis dalam bentuk matrik kinerja, sehingga akan diketahui peringkat kinerja dari masing-masing algoritma dan skenario (Khizir, et al., 2021).

HASIL DAN PEMBAHASAN

Bab ini menjelaskan hasil uji coba yang telah dilakukan sesuai metode yang telah dibahas sebelumnya. Semua tahapan tersebut dilakukan dengan bantuan mesin menggunakan aplikasi *KNIME Analytic Platforms*. Statistik dataset sebelum dan sesudah proses persiapan data ditunjukkan pada tabel 3. Terdapatnya nilai

negatif pada target *generation* dikarenakan konfigurasi instalasi PLTS yang memungkinkan untuk impor daya dari sistem PLN untuk catu daya peralatan internal. Perbedaan mencolok hanya terlihat pada kolom jumlah baris data mentah sebesar 524.287 menjadi 339.363 baris setelah proses persiapan data. Hal ini sebagai akibat dari proses penyeragaman resolusi waktu baca sensor menjadi setiap menit, proses pengambilan data sesuai periode data penelitian, dan proses pembersihan outlier.

Tabel 3. Perbandingan statistik sebelum dan sesudah proses persiapan data

Variabel	Statistik data mentah			Statistik persiapan data					
	Mean	Std. dev	Row count	Min	Max	Mean	Std. dev	Data hilang	Row count
Generation	2.293	3.432	524287	-0.078	13.684	3.188	3.668	0	339363
A-Radiation	221.756	352.839	524287	0	1574.000	310.774	386.174	0	339363
A-AmbientTmp	25.146	7.298	524287	0	36.240	25.309	8.557	0	339363
A-WindSpeed	0.087	0.142	524287	0	1.290	0.114	0.155	0	339363
A-ModuleTmp	31.473	14.475	524287	0	71.513	33.876	16.558	0	339363
B-Radiation	216.896	350.009	524287	0	1621.000	305.482	384.822	0	339363
B-AmbientTmp	24.185	8.724	524287	0	36.370	24.270	9.833	0	339363
B-WindSpeed	0.086	0.139	524287	0	1.200	0.112	0.153	0	339363
B-ModuleTmp	31.033	15.488	524287	0	72.538	33.473	17.582	0	339363

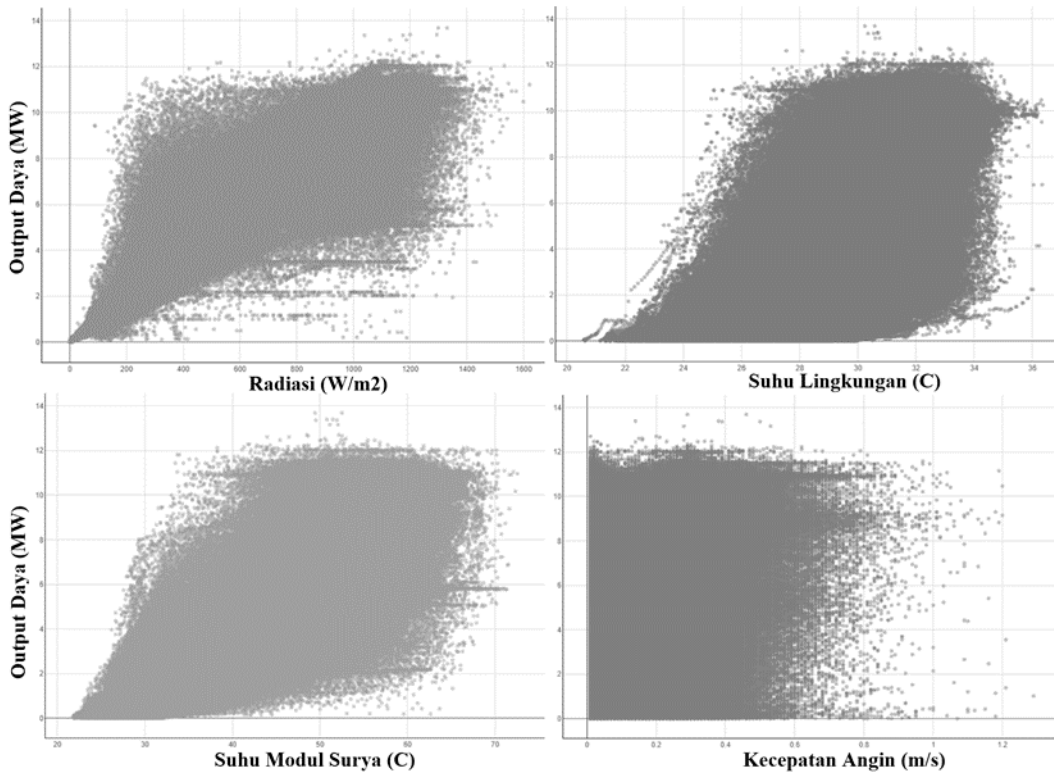
Proses selanjutnya adalah melakukan pengecekan korelasi menggunakan teknik korelasi pearson, dengan hasil ditunjukkan pada Tabel 4. Semua nilai *p-value* ≤ 0 yang berarti semua hubungan korelasi mempunyai pengaruh yang signifikan. Sehingga disimpulkan berdasarkan nilai korelasinya, produksi output daya (target) sangat dipengaruhi secara berurut oleh radiasi, suhu modul surya, kecepatan angin dan suhu lingkungan di lokasi PLTS. Berdasarkan hasil uji korelasi dan jumlah fitur yang hanya delapan item, maka diputuskan semua variable tersebut dipilih menjadi fitur dalam meramalkan keluaran daya PLTS.

Tabel 4. Nilai korelasi setiap variable Fitur dengan Keluaran Daya PLTS

Variabel Fitur	Satuan	Korelasi dengan Keluaran Daya (Target)	p-value	Kriteria Korelasi
A - Radiation	W/m ²	0.88	0,0	Sangat kuat
A - Ambient Temp	°C	0.48	0,0	Sedang/cukup
A - Wind speed	m/s	0.43	0,0	Sedang/cukup
A - Module Temp	°C	0.67	0,0	Kuat
B - Radiation	W/m ²	0.90	0,0	Sangat Kuat
B - Ambient Temp	°C	0.48	0,0	Sedang/cukup
B - Wind speed	m/s	0.46	0,0	Sedang/cukup
B - Module Temp	°C	0.70	0,0	Kuat

Data pembacaan sensor merupakan pembacaan aktual yang mewakili kondisi di lapangan. Hubungan antara keluaran daya dengan radiasi, suhu lingkungan, suhu modul surya dan kecepatan angin ditunjukkan pada gambar 4. Prinsip kerja modul surya bekerja berdasarkan besarnya radiasi yang diterima pada permukaan modul surya (Wan, et al., 2015), terlihat pada gambar 4(a) bahwa

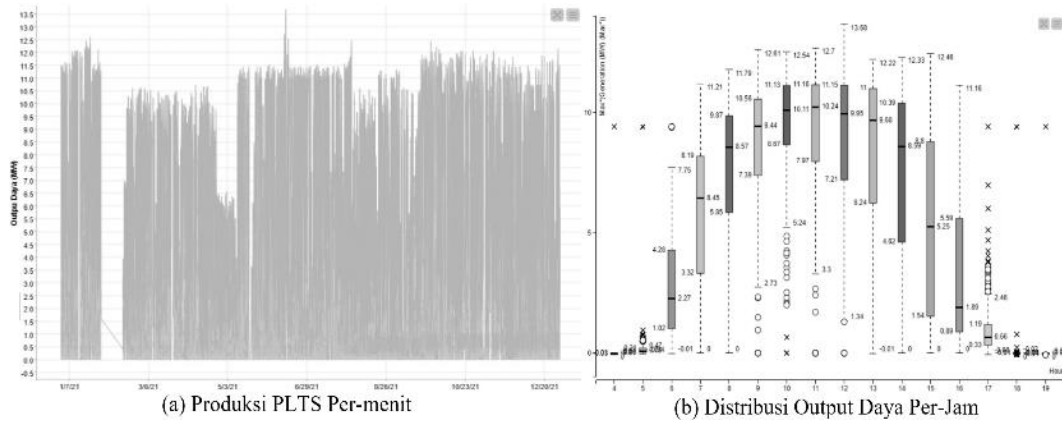
keluaran daya memiliki hubungan proporsional sangat kuat dengan radiasi. PLTS berproduksi maksimal melebihi 10 MW dengan penerimaan intensitas radiasi mulai dari 200 - 1600 W/m². Korelasi target dengan suhu lingkungan pada Gambar 4 (b) menunjukkan hubungan proporsional sedang/cukup, terutama pada kisaran suhu 26-34 ° C, menghasilkan daya keluaran maksimum melbihi 10 MW. Di sisi lain, suhu panel surya memiliki hubungan proporsional kuat pada rentang suhu 35-70 °C seperti terlihat pada gambar 4(c). Hubungan dengan variabel kecepatan angin ditunjukkan pada Gambar 4 (d), untuk nilai daya dari 0-14 MW terdistribusi pada seluruh rentang nilai kecepatan angin dari 0 hingga 1,2 m/s. Terlihat bahwa kecepatan angin bukan menjadi variabel utama penentu yang mempengaruhi produksi daya output dari PLTS, apabila dibandingkan dengan variabel lain seperti radiasi, suhu modul surya, dan suhu lingkungan sekitar.



Gambar 4. Hubungan daya keluaran (target) dengan data sensor (fitur)

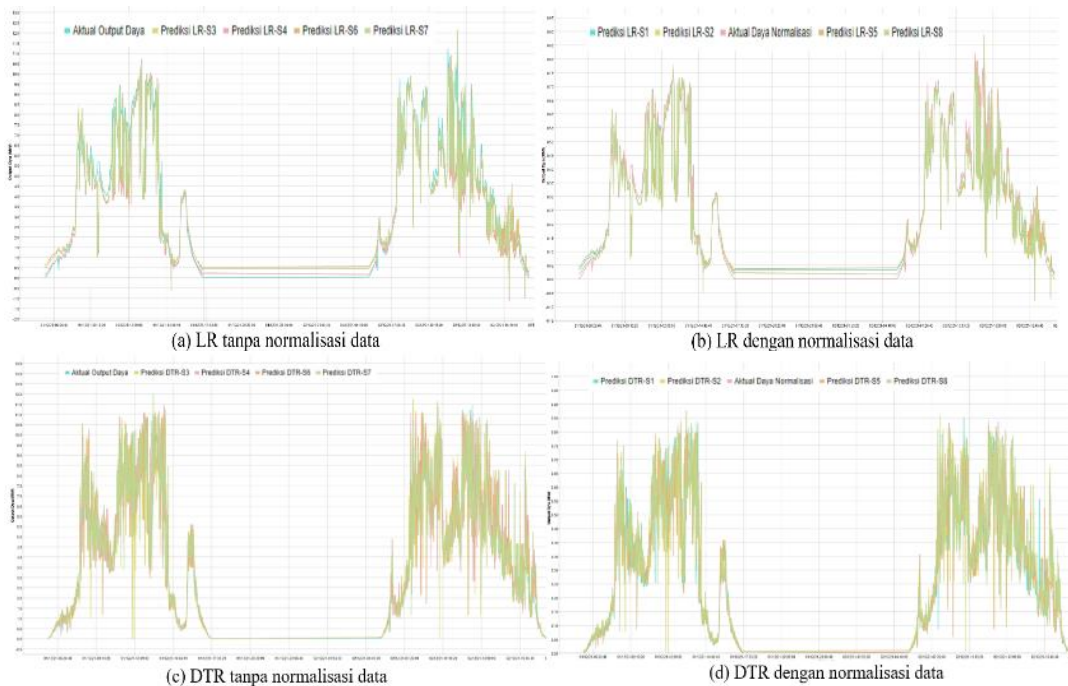
Visualisasi pada Gambar 5 menggunakan dataset sesudah dilakukan proses persiapan data. Seperti terlihat pada Gambar 5 (a), hasil pengamatan daya keluaran (target) Januari-Desember 2021 menunjukkan bahwa PLTS beroperasi sepanjang tahun 2021 dan menjalani pemeliharaan satu kali pada bulan Februari selama dua minggu. PLTS selalu berkontribusi pada jaringan kelistrikan dengan selalu memproduksi daya tiap harinya diluar jadwal pemeliharaannya, hal ini dipengaruhi oleh cuaca dan letak geografisnya yang memiliki potensi radiasi matahari sepanjang tahun. Karakteristik produksi harian ditunjukkan pada gambar 5(b), setiap hari PLTS menghasilkan listrik dari pukul 05:00-18:00 WITA, dengan pembangkitan listrik maksimum terjadi pada pukul 11.00-12.00 WITA. Fluktuasi produksi terlihat sangat besar, menunjukkan dipengaruhi secara signifikan oleh kondisi cuaca.

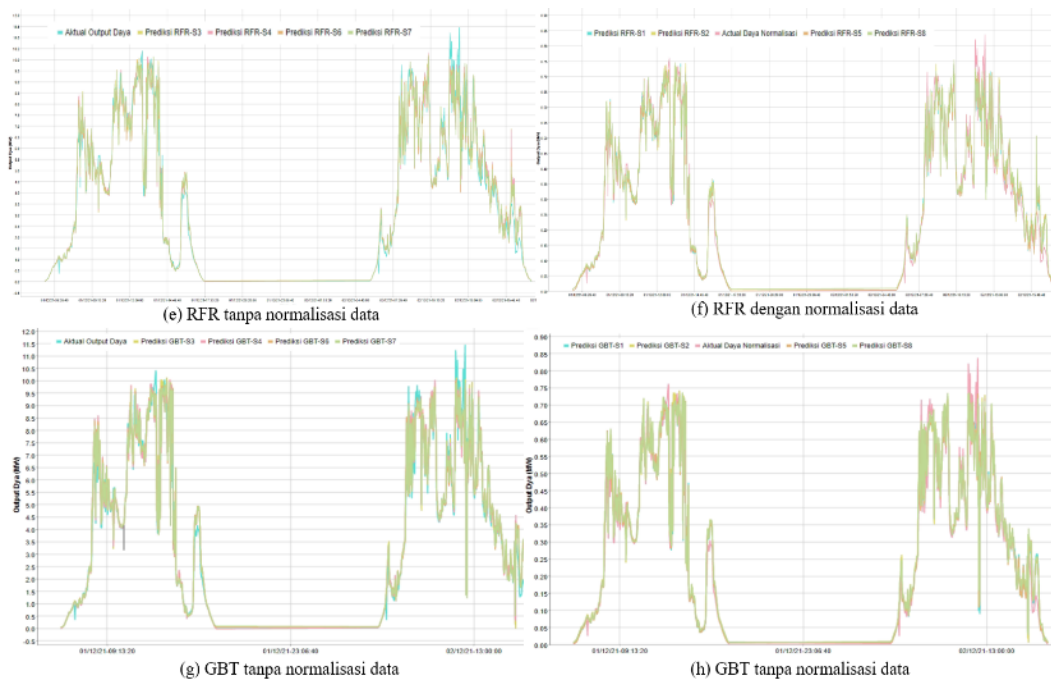
Selama tahun 2021, produksi tertinggi tercatat sebesar 13,68 MW, lebih dari 90% kapasitas terpasangnya sebesar 15 MW, dapat diasumsikan bahwa PLTS masih dalam kondisi yang baik yang ditunjukkan oleh keluaran daya tertingginya.



Gambar 5. Karakteristik Keluaran Daya Periode Januari-Desember 2021

Kinerja prediksi setiap model algoritma beserta skenarionya ditunjukkan pada Gambar 6. Grafik tersebut hanya menggunakan data sampel acak selama dua hari, karena jumlah titik data dalam resolusi waktu per menit terlalu banyak untuk ditampilkan dalam satu grafik. Proses normalisasi data merubah nilai rentan data menjadi antara 0 s.d. 1, maka grafis setiap algoritma dibagi menjadi dua bagian yaitu skenario tanpa proses normalisasi data dan skenario dengan proses normalisasi data sehingga evaluasi hasil prediksi dengan nilai aktual menjadi sebanding proporsional. Setiap algoritma menampilkan delapan skenario (S1 s.d. S8) yang sudah dijelaskan pada bab sebelumnya, dievaluasi melalui pengelompokan berdasarkan ada tidaknya proses normalisasi.





Gambar 6. Kinerja peramalan dengan berbagai macam skenario

Grafik kinerja algoritma LR dalam berbagai skenario dengan dan tanpa proses normalisasi data ditunjukkan pada gambar 6a, 6b. Terlihat bahwa pada algoritma LR tanpa normalisasi data, kinerja akurasi tertinggi dengan simpangan garis error terkecil bila dibandingkan dengan daya aktualnya adalah skenario LR-S3 dan yang terendah skenario LR-S6. Sedangkan skenario dengan normalisasi data, kinerja dengan akurasi tertinggi adalah skenario LR-S2 dan yang terendah adalah skenario LR-S5. Pada gambar 6c dan 6d dapat dilihat kinerja terbaik algoritma DTR dengan proses tanpa normalisasi data adalah skenario DTR-S4 dan yang terburuk skenario DTR-S6. Sedangkan skenario dengan normalisasi data, kinerja terbaik adalah skenario DTR-S8 dan yang terburuk skenario DTR-S5. Algoritma RFR menunjukkan kinerja terbaiknya pada skenario RFR-S3 dan akurasi terendahnya pada skenario RFR-S6 dalam proses tanpa normalisasi data (gambar 6e), sedangkan untuk data dengan normalisasi kinerja terbaiknya pada skenario RFR-S2 dan akurasi terendah pada skenario RFR-S5 (gambar 6f). Pada gambar 6g dan 6h, kinerja terbaik algoritma GBT dengan proses tanpa normalisasi data adalah skenario GBT-S7 dan dengan normalisasi data adalah skenario GBT-S1.

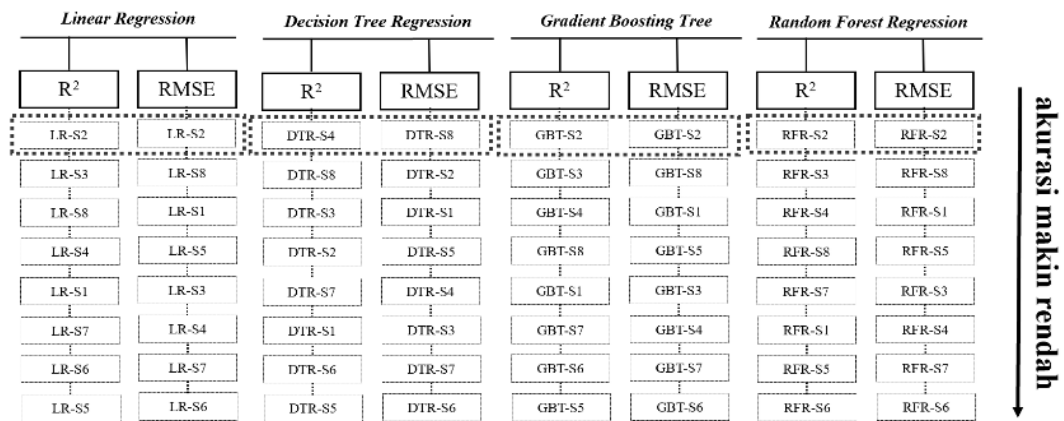
Pengamatan menunjukkan bahwa akurasi RFR lebih baik daripada LR dan DTR, namun selisih sedikit dengan GBT. LR kurang presisi dalam memprediksi daya rendah dan DTR kurang presisi dalam memprediksi daya tinggi. LR lebih akurat daripada DTR dalam memprediksi daya maksimum (puncak), tetapi memiliki kinerja akurasi terendah saat memprediksi daya minimum. Hal ini berkebalikan dengan DTR, yang akurat dalam memprediksi nilai rendah dan kurang baik dalam memprediksi nilai tinggi. LR dan DTR mempunyai error yang tinggi dalam memprediksi nilai dengan fluktuasi yang ekstrem. Pada gambar 6e,6f,6g dan 6h terlihat RFR dan GBT mempunyai akurasi yang stabil pada semua kondisi peramalan keluaran daya, baik maksimal (puncak), minimal (rendah) ataupun pada

kondisi fluktuasi yang ekstrem dibandingkan dengan LR dan DTR. Namun RFR mempunyai akurasi sedikit lebih baik pada keluaran daya fluktuatif dibanding GBT. Tabel 5 melengkapi analisis grafis, menunjukkan matrik kinerja akurasi dari hasil prediksi setiap algoritma yang dikelompokkan berdasarkan jenis algoritma beserta skenario dan indikator perhitungan akurasinya seperti R^2 dan RMSE.

Tabel 5. indikator kinerja algoritma dengan berbagai skenario

Sn	LR		DTR		GBT		RFR	
	R2	RMSE	R2	RMSE	R2	RMSE	R2	RMSE
S1	0.8718	0.0907	0.8619	0.0941	0.9461	0.0588	0.9444	0.0597
S2	0.9190	0.0695	0.9205	0.0689	0.9679	0.0438	0.9679	0.0438
S3	0.9190	0.9570	0.9209	0.9459	0.9679	0.6026	0.9678	0.6033
S4	0.9140	0.9900	0.9226	0.9393	0.9668	0.6149	0.9677	0.6067
S5	0.8687	0.0918	0.8520	0.0975	0.9417	0.0612	0.9410	0.0616
S6	0.8689	1.2565	0.8528	1.3310	0.9431	0.8276	0.9406	0.8453
S7	0.8718	1.2406	0.8625	1.2847	0.9461	0.8042	0.9445	0.8161
S8	0.9140	0.0719	0.9225	0.0682	0.9668	0.0447	0.9676	0.0441

Pada algoritma LR, S2 menjadi skenario dengan hasil akurasi terbaik baik indikator R^2 ataupun RMSE, jika diurutkan nilai skor R^2 dari yang tertinggi sampai terendah, pengaturan parameter yang paling berpengaruh terhadap hasil akurasi secara berurut adalah adanya proses perbaikan hilang rekam dan dengan adanya proses filter nilai produksi, sedangkan ada tidaknya proses normalisasi data bersifat acak atau mempunyai pengaruh yang kecil terhadap hasil akurasi yang dihasilkan oleh model. Melihat semua skor indikator akurasi untuk model algoritma LR, hasil akurasi terbaik ditujukan oleh skenario S2 yaitu melalui perbaikan data lompat, normalisasi data dan tanpa filter produksi. Algoritma DTR mempunyai akurasi terbaik pada skenario S4 untuk nilai skor R^2 dan S8 pada indikator RMSE. Urutan pengaruh pengaturan parameter yang berpengaruh terhadap akurasi adalah tanpa filter produksi, dengan normalisasi data dan tanpa perbaikan hilang rekam. Pengamatan akurasi model algoritma RFR dan GBT menunjukkan performa terbaik pada S2 untuk indikator R^2 dan RMSE. Dengan kontribusi pengaruh pengaturan parameter yang paling besar secara berurut adalah tanpa proses filter produksi, adanya proses perbaikan hilang rekam dan dengan proses normalisasi data.

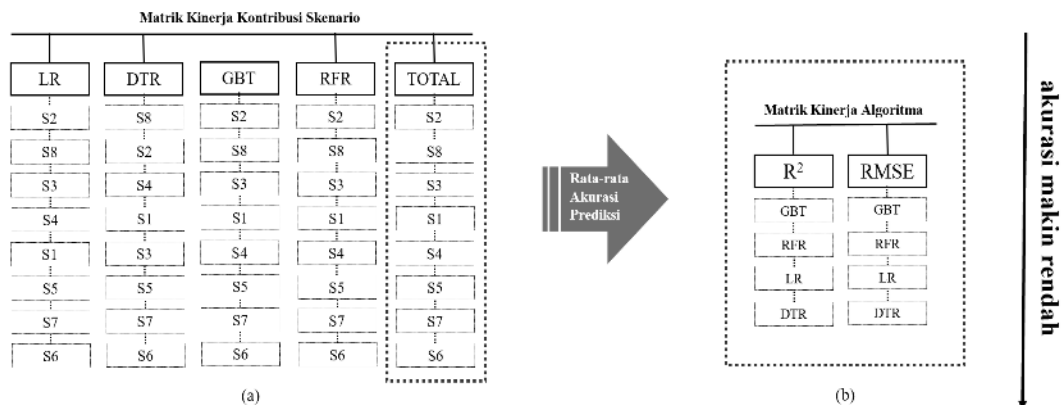


Gambar 7. Urutan kinerja kombinasi algoritma dan skenario

Nilai indikator akurasi kinerja hasil peramalan setiap algoritma dengan berbagai skenario yang tercantum dalam Tabel 5, selanjutnya diurutkan akurasinya sesuai masing-masing indikator berdasarkan algoritmanya. Gambar 7 menunjukkan kinerja skenario pada tiap algoritma berdasarkan indikator yang dipakai, urutan dari atas kebawah berarti akurasi paling tinggi dan semakin kebawah akurasinya akan melemah. Analisa lebih jauh dilakukan untuk mengetahui kontribusi pemilihan skenario terhadap akurasi hasil peramalan tiap algoritma, langkah pertama dilakukan pembobotan poin pada masing-masing skenario sesuai urutan skornya pada tiap indikator. Kemudian akan diperoleh total bobot poin tiap algoritma untuk mengetahui pengaruh kontribusi skenarionya, juga total bobot secara keseluruhan untuk mengetahui kontribusi pemilihan skenario terhadap keempat algoritma yang dipakai dalam penelitian. Simulasi pembobotan bisa dilihat pada tabel 6.

Tabel 6. simulasi pembobotan skenario terhadap nilai indikator akurasi

S _n	LR			DTR			GBT			RFR			Keseluruhan		
	R ²	RMSE	Tot	R ²	RMSE	Tot	R ²	RMSE	Tot	R ²	RMSE	Tot	R ²	RMSE	Tot
S1	4	6	10	3	6	9	4	6	10	3	8	11	14	26	40
S2	8	8	16	5	7	12	8	8	16	8	5	13	29	28	57
S3	7	4	11	6	3	9	7	4	11	7	4	11	27	15	42
S4	5	3	8	8	4	12	6	3	9	6	3	9	25	13	38
S5	1	5	6	1	5	6	1	5	6	2	5	7	5	20	25
S6	2	1	3	2	1	3	2	1	3	1	1	2	7	4	11
S7	3	2	5	4	2	6	3	2	5	4	2	6	14	8	22
S8	6	7	13	7	8	15	5	7	12	5	7	12	23	29	52



Gambar 8. Matrik kinerja akurasi permalah masing-masing algoritma

Hasil pembobotan untuk kombinasi tiap algoritma beserta skenario dengan bobot performa akurasi terbaik adalah LR-S2, DTR-S8, GBT-S2, RFR-S2. Dapat dilihat pada gambar 8a, secara keseluruhan skenario pengaturan parameter dengan hasil akurasi tertinggi secara berurutan adalah S2, S8, S3, S1, S4, S5, S7, dan S6. Analisa untuk mengetahui kinerja algoritma terbaik, dilakukan dengan cara melakukan pengelompokan nilai skor semua hasil simulasi yang dikelompokkan sesuai jenis indikator akurasinya. Setiap algoritma akan dicari nilai rata-rata skor indikatornya dari semua skor yang didapat dari kombinasi berbagai skenario. Dapat dilihat bahwa R² dan RMSE mempunyai 32 nilai skor, semua nilai tersebut kemudian dicari nilai rata-ratanya untuk masing-masing algoritma dan jenis indikatornya. Proses pengurutan dan pengelompokan tersebut beserta hasil kinerja algoritma terbaik ditunjukkan pada gambar 8b.

Analisa kinerja algoritma terbaik untuk dataset yang diuji dengan mengambil nilai rata-rata dari semua hasil prediksi dari berbagai skenario menunjukkan bahwa algoritma GBT memberikan akurasi terbaik untuk semua indikator akurasi R² dan RMSE, DTR memiliki kinerja terendah untuk indikator R² dan RMSE. Namun dalam beberapa pengamatan yang ditunjukkan oleh tabel 5, kombinasi DTR dan pilihan skenario yang tepat, mampu memberikan kinerja yang lebih baik mendekati RFR dan melampaui LR yaitu pada DTR-S4, DTR-S8, DTR-S3, DTR-S2. Sedangkan algoritma LR sebagai dasar regresi cenderung berada di kinerja pertengahan antara RFR dan DTR. Kinerja terbaik kombinasi algoritma dan skenario adalah GBT-S2 dan RFR-S2 dengan nilai yang sama pada kedua indikator yaitu 0.9679 untuk skor R² dan 0.0438 untuk skor RMSE.

KESIMPULAN

Tujuan utama dari penelitian ini adalah menemukan model peramalan jangka pendek produksi keluaran daya PLTS di sistem kelistrikan Sulutgo berbasis pembelajaran mesin melalui analisa komparatif terhadap algoritma Random Forest Regression (RFR), Gradient Boosting Trees (GTR), Linear Regression (LR) dan Decision Tree Regression (DTR) berdasarkan hasil prediksi dengan akurasi tertinggi. Dalam proses pemodelan dilakukan skenario pengaturan beberapa parameter seperti proses perbaikan hilang rekam, normalisasi data dan filter produksi. Sehingga ada delapan skenario yang dikombinasikan dengan empat algoritma di atas, yang kemudian dilakukan analisa terperinci hasil peramalan dari

beberapa matrik kinerja dari nilai R-Squared (R^2) dan RMSE. Telah dilakukan juga analisa korelasi parameter cuaca dari pembacaan sensor menunjukkan bahwa radiasi, kecepatan angin, dan suhu peralatan mempunyai korelasi yang kuat dan signifikan mempengaruhi output daya PLTS, sedangkan suhu lingkungan memiliki korelasi level sedang cenderung lemah. Analisa matrik kinerja menunjukkan bahwa secara keseluruhan GBT menjadi algoritma terbaik dengan rata-rata kinerja tertinggi dari berbagai skenario, lebih baik 0.06% dari RFR, 6.99% dari LR dan 7.46% dari DTR. Namun apabila melihat nilai akurasi tertinggi GBT dan RFR mempunyai kinerja terbaik dengan skor yang sama yaitu $R^2=0.9679$ dan $RMSE= 0.0438$. Dari perbandingan nilai R^2 tertingginya, menunjukkan kinerja GBT dan RFR lebih baik RFR lebih baik 4,91% dari DTR dan 5,32% dari LR. Pemilihan skenario yang tepat pada tiap algoritma terbukti memberi peningkatan kinerja akurasi peramalan setiap algoritma sebesar 2.78% pada algoritma GBT, 2,90% pada algoritma RFR, 8.28% pada algoritma DTR dan 5,79% pada algoritma LR yang didapatkan dari selisih kinerja terendah dan tertinggi tiap algoritma.

Dalam peramalan jangka pendek produksi PLTS sangat dipengaruhi oleh parameter selain yang telah digunakan pada penelitian ini seperti kondisi atmosfer, kondisi cuaca, pergerakan awan, spesifikasi dan konfigurasi dari panel surya yang digunakan, faktor internal dan eksternal lainnya. Namun dalam penelitian ini tidak bisa diverifikasi kinerja algoritma beserta skenarionya terhadap variabel tersebut dikarenakan data tidak tersedia. Selain itu, peningkatan akurasi hasil peramalan dari model akan sangat berpengaruh dalam pengelolaan operasi ekonomis jaringan, karena operator jaringan akan bisa lebih presisi melakukan pengaturan unit pembangkit dan penjadwalan ekonomis (*economic dispatch*) dalam tahap perencanaan dan operasi *realtime* sistem tenaga listrik. Sehingga memberi kontribusi efisiensi bisnis terutama manfaat finansial kepada operator pengelola jaringan transmisi dalam hal biaya pokok penyediaan (BPP) listrik, namun dalam penelitian ini belum dilakukan analisa kajian finansial tersebut dikarenakan data pendukung belum tersedia.

SARAN

Pada penelitian selanjutnya, perlu dilakukan penambahan data fitur yang belum dibahas dalam penelitian ini sehingga peramalan output daya PLTS bisa dilakukan lebih komprehensif. Selain itu model peramalan juga bisa dikembangkan lagi dengan algoritma yang lebih canggih seperti *deep learning* atau *extreme learning machine*. Sehingga lebih mudah untuk melakukan penyesuaian dalam model (custom) untuk mendapatkan hasil yang lebih akurat, skenario penelitian juga bisa dikembangkan lagi seperti melakukan pengelompokan cuaca, lokasi dan faktor lainnya sesuai kebutuhan operator jaringan dalam bisnis penyediaan energi listrik. Agar model peramalan bisa langsung digunakan oleh operator jaringan (*dispatcher*), perlu dikembangkan integrasi model dengan sistem informasi beserta visualisasinya yang berfungsi sebagai *dashboard* panduan operator jaringan dalam perusahaan energi listrik secara *realtime*.

Peningkatan akurasi hasil peramalan dari model akan menjadi isu strategis kepada manajemen perusahaan penyedia listrik, karena mendapatkan manfaat finansial berupa penghematan BPP listrik. Penghematan biaya bisa didapat dari

perhitungan deviasi atas peramalan terhadap produksi PLTS, apabila PLTS berproduksi lebih tinggi dari yang diperkirakan sehingga membuat frekuensi sistem naik di atas 50 Hz, kerugian akibat kelebihan frekuensi tersebut bisa dikonversi menjadi rupiah dengan mengalikan terhadap *variable cost* (Rp/Kwh) dari pembangkit base load sebagai follower dari pengoperasian PLTS.

DAFTAR RUJUKAN

- Abinet, T. E., Jianhua, Z. & Dehua, Z., 2018. Short-term photovoltaic solar power forecasting using a hybrid Wavelet-PSO-SVM model based on SCADA and Meteorological. *Renewable Energy*, Osa/vuosikerta 118, pp. 357-367.
- Budholiya, K., Shrivastava, S. K. & Sharma, V., 2020. An optimized XGBoost based diagnostic system for effective prediction of heart disease. *Journal of King Saud University - Computer and Information Sciences*, 10.
- Cebucean, D., V, C. & I, I., 2020. Modeling and performance analysis of subcritical and supercritical coal-fired power plants with biomass co-firing and CO2 capture. *Clean Technol. Environ. Policy*, Osa/vuosikerta 22, p. no. 1.
- Chen, T. & Guestrin, C., 2016. XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 8, Osa/vuosikerta 13-17-August-2016, pp. 785-794.
- Csereklyei, Z., S, Q. & T, A., 2019. The Effect Of Wind And Solar Power Generation On Wholesale Electricity Prices In Australia. *Energy Policy*, Osa/vuosikerta 131, pp. 358-369.
- Dimitris, L. ym., 2014. *News Articles Classification Using Random Forests and Weighted Multimodal Features*. Switzerland, Springer, Cham.
- Khizir, M. ym., 2021. Machine Learning Based PV Power Generation Forecasting in Alice Spring. *IEEE Access*, Osa/vuosikerta 9, pp. 46117-46128.
- Leva, S. ym., 2017. Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power. *Math Comput Simul*, Osa/vuosikerta 131, pp. 88-100.
- Menteri ESDM, R. & PLN, 2021. *Rencana Usaha Penyediaan Tenaga Listrik PT PLN 2021-2030*. Jakarta: Kementerian Energi dan Sumber Daya Mineral.
- Mohy-ud-din, G., K, M. M. & D, S., Juni 2019. Transactive energy based planning framework for VPPs in a co-optimised day-ahead and real-time energy market with ancillary services. *IET Gener., Transmiss. Distrib*, Osa/vuosikerta 13 no 11, pp. 2024-2035.
- Monowar, H., Saad, M., Malihe, D. & Lanre, O., 2017. Application of extreme learning machine for short term output power forecasting of three grid-connected PV systems. *Cleaner Production*, Osa/vuosikerta 167, pp. 395-405.
- Müller, A. C. & Guido, S., 2016. *Introduction to Machine Learning with Python*. s.l.:O'Reilly Media, Inc..
- Sobrina, S., Sam, K.-K. & Nasrudin Abd, R., 2018. Solar Photovoltaic Generation Forecasting Methods: A Review. *Energy Conversion and Management*, Osa/vuosikerta 156, pp. 459-497.

- Sobri, S., S, K. K. & N A, R., 2016. Effective Utilization Of Available PEV Battery Capacity For Mitigation Of Solar PV Impact And Grid Support With Integrated V2G Functionality. *IEEE Trans Smart Grid*, Osa/vuosikerta 7 no 3.
- Wan, C. ym., 2015. Photovoltaic and solar power forecasting for smart grid energy management. *CSEE J Power Energy Syst*, Osa/vuosikerta 1, pp. 38-46.
- Yeager, K., 2022. *SPSS Tutorials: Pearson Correlation*. [Online] Available at: <https://libguides.library.kent.edu/spss/pearsoncorr> [Haettu 10 June 2022].