



# Forecasting Produksi Energi Photovoltaic Menggunakan Algoritma Random Forest Classification

**Ergi Putra Febtiawan\*, Lalu A. Syamsul Irfan Akbar, A Sjamsjiar Rachman**

Fakultas Teknik, Program Studi S1 Teknik Elektro, Universitas Mataram, Mataram

Jl. Majapahit No.62, Gomong, Kec. Selaparang, Kota Mataram, Nusa Tenggara Barat, Indonesia

Email: <sup>1,\*</sup>ergiputraf@gmail.com, <sup>2</sup>irfan@unram.ac.id, <sup>3</sup>asrachman@unram.ac.id,

Email Penulis Korespondensi: ergiputraf@gmail.com

Submitted: 08/07/2024; Accepted: 17/07/2024; Published: 24/07/2024

**Abstrak**—Energi surya merupakan salah satu sumber energi paling terjangkau dan sebagai sumber energi terbarukan yang bersih dan ramah di dunia. Energi surya yang melimpah diseluruh dunia dapat menjadi sumber energi yang ekonomis dalam penerapannya. Disebabkan kelebihan yang dimilikinya, pemanfaatan energi surya terus meningkat sepanjang tahun. Untuk mengkonversi energi surya menjadi energi listrik salah satunya adalah dengan menggunakan perangkat Photovoltaic (PV) atau PV Module. Modul PV yang dijadikan studi terletak PLTS PLN 7 MWp desa sengkol. Peramalan produksi energi PV dilakukan untuk membantu dalam mengelola perencanaan dan mengetahui produksi energi PV berdasarkan data meteorologis pada suatu daerah yang belum terpasang perangkat PV dengan interface user berupa website. Machine learning merupakan teknik yang memungkinkan mesin belajar langsung dari contoh, data, dan pengalaman. Berbeda dengan pendekatan pemrograman tradisional yang mana mesin diberikan perintah satu persatu, machine learning dapat membuat keputusan sendiri terhadap suatu permasalahan setelah ia "belajar" dari contoh atau data yang diberikan. Penelitian ini berupaya untuk melakukan peramalan dari produksi energi yang dihasilkan perangkat photovoltaic dengan menggunakan algoritma Random Forest Classification yang mana merupakan algoritma klasifikasi dari machine learning. Algoritma ini bekerja dengan mengumpulkan prediksi dari sejumlah besar pohon keputusan yang independen, dan kemudian menggabungkan prediksi ini untuk menghasilkan hasil akhir yang lebih akurat dan stabil dengan melibatkan penggunaan data historis tentang suhu, peninjauan matahari, radiasi matahari, dan karakteristik lingkungan lainnya untuk memprediksi jumlah energi yang dihasilkan oleh sistem fotovoltaik pada waktu tertentu, sehingga didapatkan hasil kinerja keseluruhan permodelan ini terhadap data yang diberikan. Model yang dibangun dievaluasi menggunakan confusion matrix dan hasil menunjukkan bahwa algoritma ini mencapai tingkat akurasi sebesar 96%. Hasil ini mengindikasikan bahwa Random Forest Classification merupakan metode yang efektif dan andal untuk peramalan produksi energi fotovoltaik.

**Kata Kunci:** Energi Surya; Random Forest Classification; Machine Learning; Confusion Matrix; Website

**Abstract**—Solar energy is one of the most affordable energy sources and a clean and friendly renewable energy source in the world. Solar energy, which is abundant throughout the world, can be an energy source that is economical to implement. Due to its advantages, the use of solar energy continues to increase throughout the year. One way to convert solar energy into electrical energy is to use a Photovoltaic (PV) or PV Module device. The PV module used as a study is located in the PLN PLTS 7 MWp in Sengkol village. The PV module used as a study is located in the PLN PLTS 7 MWp in Sengkol village. PV energy production forecasting is carried out to assist in managing planning and knowing PV energy production based on meteorological data in an area where PV equipment has not been installed with a user interface in the form of a website. Machine learning is a technique that allows machines to learn directly from examples, data and experience. In contrast to traditional programming approaches where machines are given commands one by one, machine learning can make its own decisions on a problem after it "learns" from the examples or data provided. This research attempts to forecast the energy production produced by photovoltaic devices using the Random Forest Classification algorithm, which is a classification algorithm from machine learning, so that the overall performance results of this modeling are obtained for the data provided. This algorithm works by collecting predictions from a large number of independent decision trees, and then combining these predictions to produce a more accurate and stable final result by involving the use of historical data about temperature, solar radiation, solar radiation, and other environmental characteristics to predict the amount of energy produced by the photovoltaic system at a certain time, so that the overall performance results of this model are obtained for the data provided. The model built was evaluated using a confusion matrix and the results showed that this algorithm achieved an accuracy level of 96%. These results indicate that Random Forest Classification is an effective and reliable method for forecasting photovoltaic energy production.

**Keywords:** Solar Energy; Random Forest Classification; Machine Learning; Confusion Matrix; Website

## 1. PENDAHULUAN

Energi listrik menjadi perhatian penting di semua negara. Kehidupan manusia dan gaya hidup di zaman modern memiliki hubungan yang sangat erat dengan ketersediaan energi dan kualitasnya. Di Indonesia, berdasarkan data dari Kementerian Energi dan Sumber Daya Mineral (ESDM) Republik Indonesia, menyatakan bahwa konsumsi listrik per kapita nasional pada tahun 2019 mengalami peningkatan sebesar 2,26 % dari tahun sebelumnya dengan kenaikan mencapai 1.084 kWh per kapita. Sedangkan, pada tahun 2020 mengalami peningkatan sebesar 1.089 kWh per kapita dan pada kuartal III tahun 2021 mengalami peningkatan sebesar 1.109 kWh per kapita. Nilai ini setara dengan 92,22 % dari target yang ditetapkan pada tahun 2021 yakni sebesar 1.203 kWh per kapita. Sehingga diperlukan tindakan peramalan untuk menentukan potensi energi listrik terbarukan di berbagai daerah dengan data yang ada, serta melakukan pengujian terhadap model peramalan yang digunakan. Pembangkit Listrik Tenaga Surya atau PLTS adalah sebuah sistem yang digunakan untuk mengubah energi cahaya matahari menjadi energi



listrik dengan menggunakan prinsip efek photovoltaic. Photovoltaic sendiri merupakan fenomena fisika yang terjadi pada permukaan sel surya (solar cell) ketika menerima cahaya matahari. Selanjutnya, cahaya yang diterima diubah menjadi energi listrik. Penelitian ini diharapkan dapat membantu pengembangan PLTS di masa depan sehingga diperlukan metode peralaman yang kuat dengan metode random forest classification untuk menentukan energi listrik yang akan dihasilkan yang berguna untuk masyarakat sekitar [1].

Machine learning merupakan sub dari bidang keilmuan kecerdasan buatan (Artificial intelligence) yang banyak diteliti dan digunakan untuk memecahkan berbagai masalah[2]. Dalam implementasinya, machine learning memiliki banyak metode yang dapat digunakan untuk menangani klasifikasi, clustering, dan pengelolaan data lainnya. Dalam penelitian, digunakan beberapa model machine learning yang diantaranya terdapat algoritma Random Forest [3]. Beberapa penelitian terdahulu menggunakan algoritma random forest untuk mempelajari hubungan antar konsep dari keterkaitan yang diekstrak dan menganalisis beberapa atribut yang menyebabkan pengaruh dari data keluaran yang dihasilkan. Algoritma Random Forest banyak memberikan kontribusi terhadap berbagai penelitian [4].

Penelitian yang dilakukan oleh Harshit dkk. adalah dampak penerapan kendaraan listrik dalam skala besar yang menyebabkan kelebihan beban pada jaringan listrik karena tidak terjadwal pengisian daya EV, untuk mengurangi kelebihan beban ini, algoritma penjadwalan grid yang tepat diperlukan untuk menjadwalkan pengisian daya kendaraan listrik dan peningkatan pengisian daya publik permintaan EV. Data dapat dimanfaatkan memprediksi berbagai parameter seperti konsumsi energi, waktu pengisian daya, dengan menggunakan berbagai pembelajaran mesin (ML) populer seperti random forest yang membahas evaluasi model dengan root mean square error (RMSE) dan mean absolute percentage error (MAPE). Hasil untuk ketiga model yang digunakan, dengan algoritma RF yang merupakan jenis teknik ensemble untuk memvisualisasikan kepentingan variabel dan untuk tujuan prediksi ini untuk konsumsi energi didapatkan skor RMSE 5,9% dan MAPE 0,188, dari XGBoost didapatkan RMSE untuk konsumsi energi 9% dan MAPE 0,288 dan dari regresi linier didapatkan RMSE untuk konsumsi energi sebesar 10,22% dan MAPE sebesar 0,349 [5].

Penelitian yang dilakukan oleh Ardianto dkk. adalah 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. Data histori 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<sup>2</sup> sebesar 0.9679 dan RMSE sebesar 0.0438. Pemilihan skenario yang tepat terbukti memberi peningkatan kinerja akurasi sebesar RFR 2,90% [6].

Penelitian yang dilakukan oleh Gufron dkk. adalah Isu krisis energi menuntut orang mencari sumber energi alternatif. PLTS menjadi pilihan yang menjanjikan untuk menjawab tantangan krisis energi tersebut. Namun PLTS tergantung oleh kondisi cuaca, sangat sulit memperkirakan berapa produksi energi pada suatu sistem PLTS. Penelitian peramalan produksi energi 1kWp menggunakan mesin pembelajaran dan support vector machine (SVM) telah dilakukan dan dibandingkan dengan model multiple linear regression (MLR), model peramalan dengan pendekatan deret waktu, data training periode Januari – Desember 2021 dan data tes periode Januari – Mei 2022. Konstruksi ramalannya adalah hasil produksi tiga hari kebelakang meramalkan produksi energi hari berikutnya. Hasil evaluasi MAPE pada data training SVM dan MLR adalah 19.79% dan 23.96%, sedangkan pada testing 21,79% dan 20.45%. Hasil peramalan harian diakumulasi perbulan dan dievaluasi, hasilnya MAPE 4.13% dan 5.56% masing-masing untuk SVM dan MLR. Kedua model SVM dan MLR layak dikembangkan lebih lanjut pada forecasting PLTS 1 kWp berdasarkan data deret waktu [7].

Penelitian yang dilakukan oleh Gökçe dkk. adalah kebutuhan selama periode ketika permintaan energi mencapai puncaknya dan menghambat pasar agar tidak mengalami kerugian ekonomi pada saat berada di level terendah, prediksi terdekat harus dibuat. Penelitian ini, mempertimbangkan jumlah Kesalahan Kuadrat dari model, Model Linear menunjukkan kinerja terburuk dengan nilai RMSE 2348,35, sedangkan teknik XGBoost menunjukkan kinerja yang lebih baik dengan nilai RMSE 2038,54, sisanya di bawah nilai RMSE 2075,32 yang dihasilkan dari penerapan Model Random Forest. Oleh karena itu, model dengan kinerja prediksi tertinggi dibandingkan dengan jumlah kesalahan kuadrat ditentukan sebagai Random Forest. Ketika jumlah kesalahan persentase dibandingkan dalam fase peramalan semua model, urutan kinerja ditentukan sebagai Model Random Forest, XGBoost dan Model Linear [8].

Penelitian yang dilakukan oleh Da liu dkk. adalah Karakteristik fotovoltaik yang terputus-putus dan tidak terkendali pembangkit listrik mempunyai dampak yang sangat besar terhadap stabilitas sistem tenaga listrik. Untuk mengurangi terjadinya kondisi seperti itu, perlu dilakukan peningkatan akurasi prediksi pembangkit listrik fotovoltaik. Algoritma HGWO digunakan untuk memilih dengan cepat parameter optimal model dengan verifikasi ditetapkan sebagai sasarannya. Menggunakan MAE, RMSE dan Gambar prediktif analisis dalam percobaan, ditemukan bahwa algoritma yang direkomendasikan PCA-K-means-HGWO-RF mewarisi ketidakpekaan dan

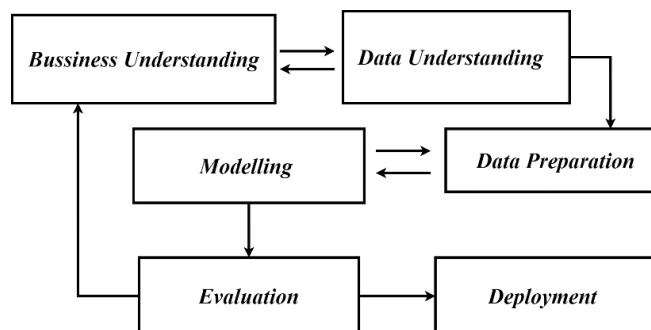
kemampuan penyesuaian Random Forest yang baik terhadap data noise, dan akurasinya dan stabilitas prediksi algoritma. Model yang direkomendasikan. Nilai MAE pada ketiga wilayah tersebut lebih rendah dari hasil optimal pada model perbandingan sebesar 0,18, 0,14 dan 0,19 pada h+1 periode. Indikator ini mencerminkan tingkat rata-rata kinerja prediktif model yang direkomendasikan dibandingkan model lainnya. Karena keterbatasan algoritma model perbandingan, stabilitasnya mungkin dipengaruhi oleh data kebisingan. Misalnya, pohon keputusan tunggal adalah rentan terhadap over-fitting, parameter jaringan saraf sulit dilakukan estimasi, dan reproduktifitasnya buruk. Model yang direkomendasikan adalah juga berada di posisi terdepan pada indikator RMSE, menunjukkan bahwa model disarankan untuk memiliki stabilitas yang lebih tinggi dalam menangani poin yang tidak normal [9].

Berdasarkan referensi dari 5 penelitian tersebut, penelitian ini akan menggunakan algoritma Random Forest Classification untuk metode pembelajaran mesin yang kuat dan sering digunakan untuk masalah prediksi, termasuk dalam konteks energi fotovoltaik. Algoritma ini bekerja dengan mengumpulkan prediksi dari sejumlah besar pohon keputusan yang independen, dan kemudian menggabungkan prediksi ini untuk menghasilkan hasil akhir yang lebih akurat dan stabil. melibatkan penggunaan data historis tentang suhu, penyinaran matahari, radiasi matahari, dan karakteristik lingkungan lainnya untuk memprediksi jumlah energi yang dihasilkan oleh sistem fotovoltaik pada waktu tertentu di masa depan.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Tahapan dalam penelitian ini menggunakan metode Cross Industry Standard Process for Data Mining (CRISP-DM) yang terdiri dari tahap business understanding yang merupakan Gambaran tentang sumber daya data yang tersedia, tahap data understanding digunakan untuk mengumpulkan data dari sumber data, mengeksplorasi dan mendeskripsikannya serta memeriksa kualitas data, tahap data preparation digunakan untuk menetapkan kriteria inklusi dan eksklusi, tahap modeling terdiri dari pemilihan teknik pemodelan, pembuatan kasus uji dan model, tahap evaluation diperiksa terhadap tujuan bisnis yang telah ditentukan, dan tahap deployment berupa laporan akhir atau komponen perangkat lunak. Gambar 1 menjelaskan gagasan utama, tugas dan keluaran dari tahapan-tahapan tersebut singkatnya, berdasarkan panduan pengguna CRISP-DM [10].



Gambar 1. Langkah-langkah Penelitian [11]

### 2.2 Business Understanding

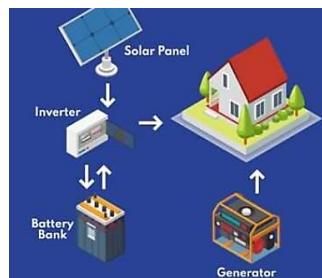
Sistem PLTS dibagi menjadi dua klasifikasi sistem. Yang pertama Gambar cadangan ketika terjadi gangguan atau pemadaman pada jaringan listrik PLN. Yang kedua yaitu PLTS berdiri sendiri atau biasa disebut dengan PLTS standalone. Sistem PLTS ini dirancang untuk beroperasi secara mandiri tanpa ada hubungan konfigurasi sistem dengan jaringan listrik PLN [1].

Teknologi yang menggunakan sel photovoltaik untuk mengubah dari sinar matahari (surya) menjadi energi listrik yang nantinya dapat langsung digunakan ke beban (self consumption) atau disebut PLTS on-grid juga untuk charging baterai dan selebihnya akan disalurkan ke dalam jaringan PLN. Alur dari PLTS on-grid dapat dilihat pada Gambar 1 [12].



Gambar 2. Alur PLTS on-grid

Sistem PLTS off-grid merupakan sistem PLTS yang tidak terhubung dengan jaringan listrik PLN. Sistem ini disebut juga System Stand Alone karena sistem ini hanya mengandalkan energi matahari sebagai satu-satunya sumber energi utama dengan menggunakan rangkaian modul surya untuk menghasilkan energi listrik sesuai dengan kebutuhan. Alur dari PLTS off-grid dapat dilihat pada Gambar 2.[13]



**Gambar 3.** Alur PLTS off-grid

### 2.3 Data Understanding

Transformasi data digunakan untuk mengubah data ke dalam bentuk yang sesuai untuk suatu analisis. Transformasi data dilakukan baik terhadap data numerik maupun data kategorik. Transformasi data numerik dilakukan dengan standarisasi data [14]. Data tersebut dapat dilihat pada Tabel 2 berikut ini.

**Tabel 1.** Deskripsi Data

Atribut	Jenis Data
Date	Numerik
X1 (PV module temperature (°C))	Numerik
X2 (In plane irradiance (W/m <sup>2</sup> ))	Numerik
X3 (Internal hygrometry (g/m <sup>3</sup> ))	Numerik
Y1 (Active energy exported (kWh))	Numerik

Hubungan antara temperature (°C) dengan Active energy exported (kWh) temperatur panel fotovoltaik secara bersama-sama berpengaruh sangat kuat terhadap daya yang dihasilkan panel fotovoltaik dengan nilai determinasi sebesar 95%. Secara parsial radiasi matahari berkorelasi positif terhadap daya output panel fotovoltaik dan memiliki nilai koefesien regresi sebesar 0,310, temperatur panel fotovoltaik berkorelasi positif terhadap daya keluaran panel fotovoltaik dan memiliki nilai koefesien regresi sebesar 0,974, dan temperatur area berkorelasi negatif terhadap daya keluaran (output) panel fotovoltaik dan memiliki nilai koefesien regresi sebesar -4,762 [15].

Hubungan antara irradiance (W/m<sup>2</sup>) dengan Active energy exported (kWh), merupakan parameter densitas (daya) sinar matahari yang diukur dalam (W/m<sup>2</sup>). Konstanta matahari bumi adalah radiasi yang diterima bumi dari matahari melalui atmosfer AM0 1367 W/m<sup>2</sup>. Setelah melewati atmosfer dengan panjang lintasan AM1, radiasi berkurang menjadi sekitar 1000 W/m<sup>2</sup> dan memiliki kandungan spektral karena penyerapan atmosfer. Penyinaran AM1.5 diterima sebagai spektrum kalibrasi standar untuk sel surya. Radiasi matahari yang masuk ke permukaan panel fotovoltaik dipengaruhi oleh kondisi cuaca, posisi matahari, dan sebagainya. Sementara itu, temperatur panel fotovoltaik dipengaruhi oleh temperatur lingkungan dan panas yang dihasilkan oleh panel saat beroperasi [15].

Hubungan antara hygrometry (g/m<sup>3</sup>) dengan Active energy exported (kWh), pengaruh kelembaban udara terhadap daya listrik, diketahui nilai signifikansi variabel kelembaban udara adalah sebesar 0,74. Karena nilai Sig. 0,74 lebih besar dari probabilitas 0,05 maka dapat disimpulkan bahwa tidak ada pengaruh yang signifikan antara kelembaban udara terhadap daya listrik. nilai hitung variabel kelembaban udara adalah 0.00625 sedangkan nilai Tabel adalah 2,2281. Karena nilai hitung lebih kecil daripada Tabel maka disimpulkan tidak ada pengaruh yang signifikan antara kelembaban udara terhadap daya listrik [16].

### 2.4 Data Preparation

Setelah proses data understanding dilakukan, berikutnya akan diberikan adalah dengan melakukan proses klasifikasi untuk penanganan missing value Memeriksa apakah terdapat nilai yang hilang pada setiap variabel (X1 hingga Y1) untuk menentukan apakah memerlukan tindakan pengisian nilai yang hilang atau langkah khusus untuk menanggulangi masalah tersebut. sebelum diproses lebih lanjut pada tahapan dan fase pemrosesan selanjutnya. [17]

Data Scaling atau normalisasi Mean Normalization efektif dalam skenario ketika distribusi data tidak diasumsikan normal dan ketika rentang nilai harus dipusatkan di sekitar nol[18]. Standarisasi suatu dataset melibatkan perubahan skala distribusi nilai sehingga rata-rata (mean) yang diamati adalah 0 dan simpangan baku adalah 1. standard deviation dihitung menggunakan di mana  $z_{\text{mean}}$  adalah rata-rata data. Normalisasi dapat dihitung dengan menggunakan dimana  $z_i$  merupakan vektor ciri asli,  $z_{\text{mean}}$  merupakan mean dari vektor ciri,  $z_{\text{std}}$  merupakan standard deviation, dan  $z_i$  merupakan nilai hasil standardisasi [19].



$$z_{\text{std}} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (z_i - z_{\text{mean}})^2} \quad (1)$$

$$z_i \left( \frac{z_i - z_{\text{mean}}}{z_{\text{std}}} \right) \quad (2)$$

## 2.5 Labeling

Data daya eksport yang diperoleh dari situs monitoring panel surya merupakan data ordinal, sehingga perlu dilakukan penglabelan untuk mempermudah proses klasifikasi. Label yang digunakan adalah “Rendah”, “Sedang”, dan “Tinggi” dengan menetapkan range tertentu untuk setiap labelnya. Ada 2 tipe dataset yang ditemui dalam masalah pengklasifikasian, yaitu dataset yang tidak seimbang (imbalanced dataset) dan dataset yang seimbang (balanced dataset). Pada penelitian ini, dataset yang disiapkan adalah imbalanced dataset. Pembagian range untuk masing-masing label pada penelitian ini berdasarkan pada dataset yang dikumpulkan, dimana nilai tertinggi yang diperoleh dikurangi nilai terendahnya kemudian dibagi 3 untuk mendapatkan range masing-masing label. Cara seperti ini juga dilakukan Bayindir, Yesibuldak, Colak, dan Genc dalam penelitiannya, dimana mereka membagi target klasifikasi menjadi 5 label berdasarkan dataset yang mereka punya [20].

## 2.6 Modeling

Random Forest adalah pengembangan dari metode Decision Tree yang menggunakan beberapa Decision Tree, dimana setiap Decision Tree telah dilakukan pelatihan menggunakan sampel individu dan setiap atribut dipecah pada pohon yang dipilih antara atribut subset yang bersifat acak. Random Forest memiliki beberapa kelebihan, yaitu dapat meningkatkan hasil akurasi jika terdapat data yang hilang, dan untuk resisting outliers, serta efisien untuk penyimpanan sebuah data [21]. Rumus dari random forest di tunjukkan pada persamaan [22].

$$\text{Gini} = 1 - \sum_{i=1}^n (p_i)^2 \quad (3)$$

Decision Tree atau pohon keputusan adalah algoritma yang memisahkan sebuah kelompok data dengan menggunakan struktur pohon. Konsep dari pohon keputusan ini adalah membuat pemisahan Data Berdasarkan kondisi yang mana kondisi tersebut dibuat dalam bentuk cabang. Algoritma decision tree mempunyai beberapa algoritma, yaitu ID3 berdasarkan nilai entropy dan CART berdasarkan nilai gini. [23] Rumus dari pencarian nilai impurity pada algoritma CART ditunjukkan pada Persamaan [22].

$$E(S) = \sum_{i=0}^n p_i \log_2 p_i \quad (4)$$

Rumus dari pencarian nilai informasi gain pada algoritma CART ditunjukkan pada Persamaan.

$$IG(Y, X) = E(Y) - E(Y|X) \quad (5)$$

## 2.7 Evaluation

Pada konsep data mining, untuk menghitung akurasi pada teknik klasifikasi dapat menggunakan metode confusion matrix yang merupakan Tabel yang sering digunakan untuk menggambarkan kinerja model klasifikasi (classifier) pada satu set data uji yang nilai sebenarnya diketahui. Ini merupakan kinerja dari suatu algoritma. Presisi atau confidence adalah proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. Recall atau sensitivity adalah proporsi kasus positif yang sebenarnya yang diprediksi positif secara benar [24].

**Tabel 2.** Model confusion matrix

Aktual	Hasil Klasifikasi	
	+	-
+	True Positive (A)	True Negative (B)
-	False Positive (C)	False Negative (D)

Akurasi merupakan jumlah prediksi yang benar dibagi dengan keseluruhan sampel data. Akurasi dapat diperoleh menggunakan persamaan (6). [25]

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Presisi merupakan rasio berapa banyak prediksi yang benar-benar positif dari semua kelas positif yang diprediksi dengan benar. Nilai presisi dapat dihitung menggunakan persamaan (7) [25].

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (7)$$

Recall atau sensitivity merupakan seberapa banyak prediksi yang benar dari semua kelas positif. Nilai recall dapat dihitung dengan menggunakan persamaan (8) [25].

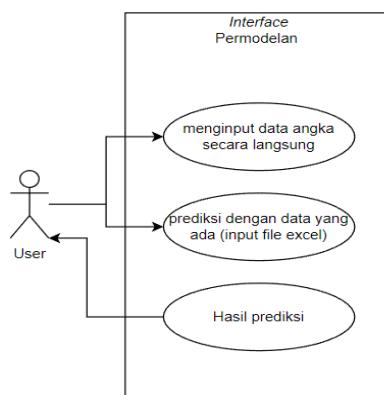
$$\text{Recall} = \frac{TP}{TP+FN} \quad (8)$$

Prediksi sampel yang banyak pada kelas positif akan mengakibatkan FP juga bertambah, dan recall bisa meningkat secara maksimal. Selanjutnya, presisi akan dibuat turun karena tujuan utama dari evaluasi model menggunakan presisi adalah mengurangi jumlah FP. Sementara itu, untuk FN yang mempengaruhi recall, berlaku juga sebaliknya. Model yang seimbang membutuhkan suatu cara. Mencari nilai tengah dari presisi dan recall yang merupakan rata-rata harmonik dari keduanya dapat menggunakan F-Score, sehingga F-Score bisa dihitung dengan menggunakan persamaan (9) [25].

$$F - \text{Score} = 2 \times \frac{(\text{Presisi} \times \text{Recall})}{\text{Presisi} + \text{Recall}} \quad (9)$$

## 2.8 Deployment

Situs web adalah kumpulan halaman dalam domain yang berisi berbagai informasi yang memungkinkan pengguna Internet untuk membaca dan melihatnya melalui mesin pencari [26]. Hasil analisis dan pemodelan yang telah dilakukan akan dibuat dalam bentuk visualisasi yang mudah dipahami. Tujuan dari tahap deployment ini adalah untuk memastikan bahwa hasil analisis dapat digunakan dan dimanfaatkan secara optimal untuk mencapai tujuan yang telah ditentukan pada tahap awal yaitu untuk melakukan peramalan keluaran energi PV yang ada di PLTS PLN Provinsi Nusa Tenggara Barat. Data yang diperoleh divisualisasikan menggunakan Website streamlite dengan library untuk Interface peramalan berdasarkan hasil permodelan.



**Gambar 3.** Use Case Diagram Interface Website

## 3. HASIL DAN PEMBAHASAN

Bagian ini menguraikan hasil yang diperoleh berdasarkan model sistem yang telah dibangun dan diimplementasikan. Beberapa tahapan pengujian dilakukan dalam penelitian ini. Salah satu tujuan dari penelitian ini adalah untuk mengetahui bagaimana kinerja dari model random forest yang dibangun dan membuat interface website yang digunakan untuk pengguna. Hasil penelitian yang dilakukan berupa hasil forecasting melalui proses dengan menggunakan algoritma random forest classification yang didukung oleh tools visual studio code dengan bahasa pemrograman python.

### 3.1 Pengolahan Data

Data dalam penelitian ini merupakan data sekunder yang diperoleh dari SCADA PLTS PLN di Desa sengkol, Provinsi Nusa Tenggara Barat, yang berupa data operasional perusahaan di tahun agustus 2023 – maret 2024. Data diperiksa apakah terdapat nilai yang hilang pada setiap variabel (X1 hingga Y1) untuk menentukan apakah memerlukan tindakan pengisian nilai yang hilang atau langkah khusus untuk menanggulangi masalah tersebut. Hal ini dilakukan agar proses modeling mendapatkan hasil yang baik berikut contoh sampel data yang di ambil dapat dilihat pada Tabel 3.

**Tabel 3.** Data Sampel SCADA

Date	PV module temperature (°C)	In plane irradiance (W/m²)	Internal hygrometry (g/m³)	Active energy exported (kWh)
08/01/2023 06:00	24.76716543	21.22462891	87.03099975	59.846
08/01/2023 07:00	24.82146612	199.1148023	97.91960182	913.961
08/01/2023 08:00	23.521614	446.9617547	97.67276995	2523.273
03/31/2024 15:00	38.33062593	336.7009837	87.74161275	1930.226
03/31/2024 16:00	31.02367437	64.73944663	92.2355561	355.351
03/31/2024 17:00	31.81075078	130.9630324	94.06727945	721.432

Setelah memilih atribut yang akan digunakan untuk proses modeling tahap selanjutnya yaitu melakukan normalisasi menjadi skala umum tanpa mengubah perbedaan rentang nilai dengan menggunakan metode standard scaling. Normalisasi data akan membantu mempercepat proses pembelajaran dalam pembelajaran mesin.

Selanjutnya memilih atribut yang akan digunakan untuk proses modelling dengan cara menghapus data yang tidak diperlukan. Atribut yang dihapus adalah data ‘Date’ dan Active energy exported (kWh), data ini dianggap tidak diperlukan karena tidak berhubungan dengan penelitian atau telah ditambahkan kolom baru sebagai representasi dari klasifikasi data tersebut. Melalui tahap data preparation ini, dilakukan proses pemilihan atribut dari dataset tersebut. Atribut yang digunakan yaitu X1 (PV module temperature (°C)), X2 (In plane irradiance (W/m<sup>2</sup>)), X3 (Internal hygrometry (g/m<sup>3</sup>)), dan Y1 (Penggunaan) sebagai kolom klasifikasi dari daya export yang sudah dihapus.

**Tabel 4.** Hasil Pemrosesan Data

PV module temperature (°C)	In plane irradiance (W/m <sup>2</sup> )	Internal hygrometry (g/m <sup>3</sup> )	Penggunaan
-1.849.500	-1.280.258	0.541273	Rendah
-1.405.550	-0.732578	1.290.761	Rendah
-0.268416	0.030483	1.273.771	Sedang
0.601633	0.693478	0.511587	Sedang
1.280.435	1.161.756	-0.504749	Tinggi

Pelabelan digunakan untuk mempermudah proses klasifikasi. Label yang digunakan adalah “Rendah”, “Sedang”, dan “Tinggi” dengan menetapkan range tertentu untuk setiap labelnya, kemudian diketahui jumlah data dari masing-masing klasifikasi dan didapatkan bahwa data yang dimiliki yaitu dataset yang tidak seimbang (imbalanced dataset).

**Tabel 5.** Label classification

Label	Range (kWh)	Jumlah
Rendah	0 > 2500.0	1.629
Sedang	2500.0 > 4000.0	781
Tinggi	4000.0 > 5000.0	824

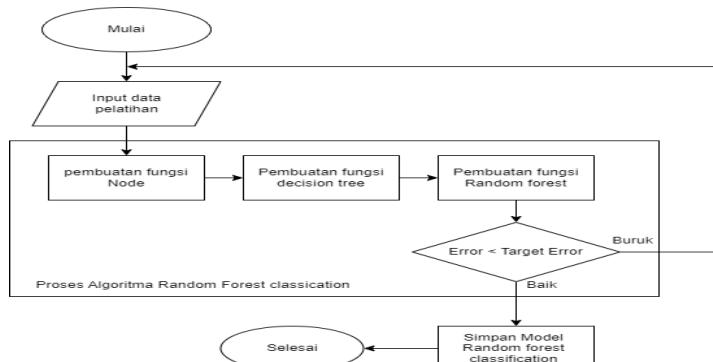
Algoritma machine learning dapat menghasilkan prediksi lebih cepat ketika diberikan format data numerik. Oleh karena itu, label encoding digunakan untuk mengubah label menjadi bentuk numerik sehingga dapat diproses lebih cepat oleh mesin. Proses encoding dilakukan dengan menggunakan bantuan program python di perangkat lunak visual studio code. Adapun label produksi energi PV yang telah dirubah menjadi bentuk numerik ditampilkan oleh Tabel 4.

**Tabel 6.** Label Encoding

Label	Hasil Encoding
Rendah	0
Sedang	1
Tinggi	2

### 3.2 Random Forest

Pemodelan random forest ini akan dibuat menggunakan data pelatihan yang telah diolah kemudian akan dimasukan ke dalam bahasa python dengan IDE Visual Studio Code. Sehingga didapatkan hasil model yang di-inginkan dengan melakukan pengujian untuk menghitung akurasi pada teknik klasifikasi dapat menggunakan metode Confusion matrix yang merupakan Tabel yang sering digunakan untuk menggambarkan kinerja model klasifikasi (classifier) pada satu set data uji yang nilai sebenarnya diketahui. Berikut langkah-langkah proses pembuatan permodelan dari algoritma random forest dapat dilihat pada Gambar 4.



**Gambar 4.** Proses Pemodelan Algoritma random forest classification

Algoritma Random Forest merupakan metode ensemble learning yang menggunakan beberapa pohon keputusan (decision tree) dan hasilnya kemudian ditentukan dari rata-rata pohon yang ada untuk dapat meningkatkan akurasi prediksi. Proses pembuatan algoritma ini dimulai dengan input data pelatihan yang akan digunakan untuk melatih model. Pertama, kita harus membuat fungsi ‘Node’ yang berfungsi sebagai komponen dasar dari pohon keputusan, yang akan menentukan aturan pemisahan data berdasarkan fitur-fitur tertentu. Kemudian, membuat fungsi untuk membangun decision tree, yang akan menggunakan fungsi Node untuk membentuk struktur pohon yang dapat digunakan untuk membuat prediksi berdasarkan data pelatihan dan terdapat perhitungan entropy atau purity serta information gain di dalamnya. Setelah itu, membuat fungsi untuk membangun random forest, yang akan menggabungkan beberapa pohon keputusan dengan cara memilih subset data dan fitur secara acak untuk setiap pohon, sehingga menghasilkan model yang lebih akurat dengan dibatasi jumlah pohon keputusan sebesar 10 sampel, kedalaman maksimum pohon keputusan sebanyak 10 sampel, dan jumlah node dalam membagi pohon keputusan sebanyak 2 sampel. Selanjutnya, kita akan menghitung persen error dari prediksi yang dihasilkan oleh model, dan memastikan bahwa error tersebut berada di bawah target error yang telah ditentukan yaitu sebesar kurang dari 5%. Jika error sudah memenuhi kriteria, langkah terakhir adalah menyimpan permodelan yang telah dibuat agar dapat digunakan untuk prediksi di masa mendatang.

Dataset yang digunakan ialah data yang telah diolah kemudian dibagi menjadi data latih dan data uji. Data latih merupakan data yang akan digunakan untuk membangun model sebesar 80% sedangkan data uji digunakan untuk menguji seberapa baik model bekerja sebesar 20%. pada satu set data uji yang nilai sebenarnya diketahui. Penelitian ini menggunakan 3 label klasifikasi sehingga dapat dilihat pada Gambar 4.

Setiap sel dalam matriks menunjukkan jumlah contoh yang termasuk dalam kategori tertentu berdasarkan prediksi model:

a.  $(0,0) = 323$

Model memprediksi 323 sampel sebagai kelas 0, dan sebenarnya sampel tersebut adalah kelas 0 (true positive untuk kelas 0).

b.  $(0,1) = 5$

Model memprediksi 5 sampel sebagai kelas 1, tetapi sebenarnya sampel tersebut adalah kelas 0 (false negative untuk kelas 0 dan false positive untuk kelas 1).

c.  $(0,2) = 0$

Model memprediksi 0 sampel sebagai kelas 2, tetapi sebenarnya sampel tersebut adalah kelas 0 (false negative untuk kelas 0 dan false positive untuk kelas 2).

d.  $(1,0) = 3$

Model memprediksi 3 sampel sebagai kelas 0, tetapi sebenarnya sampel tersebut adalah kelas 1 (false negative untuk kelas 1 dan false positive untuk kelas 0).

e.  $(1,1) = 128$

Model memprediksi 128 sampel sebagai kelas 1, dan sebenarnya sampel tersebut adalah kelas 1 (true positive untuk kelas 1).

f.  $(1,2) = 14$

Model memprediksi 14 sampel sebagai kelas 2, tetapi sebenarnya sampel tersebut adalah kelas 1 (false negative untuk kelas 1 dan false positive untuk kelas 2).

g.  $(2,0) = 0$

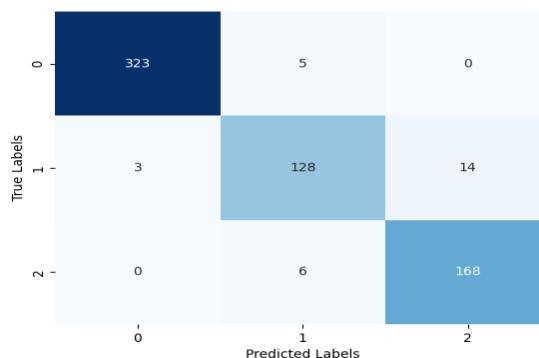
Model memprediksi 0 sampel sebagai kelas 0, tetapi sebenarnya sampel tersebut adalah kelas 2 (false negative untuk kelas 2 dan false positive untuk kelas 0).

h.  $(2,1) = 6$

Model memprediksi 6 sampel sebagai kelas 1, tetapi sebenarnya sampel tersebut adalah kelas 2 (false negative untuk kelas 2 dan false positive untuk kelas 1).

i.  $(2,2) = 168$

Model memprediksi 168 sampel sebagai kelas 2, dan sebenarnya sampel tersebut adalah kelas 2 (true positive untuk kelas 2).



**Gambar 5.** Hasil Confusion Matrix  $3 \times 3$

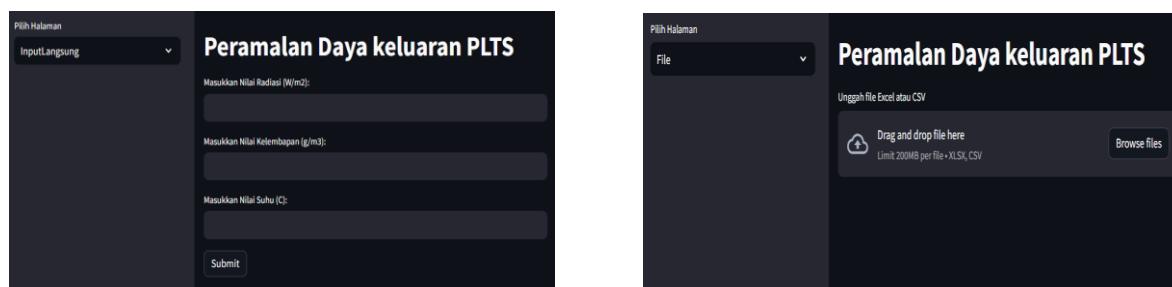
**Tabel 6.** Hasil kinerja random forest

	Presisi	recall	f1-score	Support
0	0.99	0.98	0.99	328
1	0.91	0.89	0.90	145
2	0.93	0.96	0.95	174
Accuracy			0.96	647
Macro avg	0.95	0.94	0.94	647
Weighted avg	0.96	0.96	0.96	647

Akurasi dari algoritma random forest pada data yang di tes sebesar 0.96%. Cara yang digunakan untuk mengoptimalkan model yang kurang bagus adalah dengan menambahkan data sampel. Hasil error pada model berbanding terbalik dengan akurasi untuk kasus data ini sebesar 0,04%. Kemudian untuk nilai presisi pada label “rendah” bernilai 0.99%, “sedang” bernilai 0.91%, dan “tinggi” bernilai 0,93% Presisi ini merupakan rasio berapa banyak prediksi yang benar-benar positif dari semua kelas positif yang diprediksi dengan benar. Recall atau sensitivity merupakan seberapa banyak prediksi yang benar dari semua kelas positif didapatkan pada label “rendah” bernilai 0.98%, “sedang” bernilai 0.89%, dan “tinggi” bernilai 0,96% F1-Score digunakan untuk mencari nilai tengah dari presisi dan recall yang mana merupakan rata-rata harmonik dari keduanya sehingga didapatkan nilai yang seimbang. nilai F1-Score yang didapat pada label “rendah” bernilai 0.99%, “sedang” bernilai 0.90%, dan “tinggi” bernilai 0,95%

### 3.3 Deployment

Hasil yang diperoleh divisualisasikan menggunakan Website streamlite dengan library untuk Interface peramalan berdasarkan hasil permodelan. Aktivitas utama yang dilakukan oleh user memasukkan data secara langsung, mengunggah data menggunakan file, dan mengambil serta melihat hasil input. Aktor dalam diagram ini adalah user yang berinteraksi dengan sistem melalui tiga kasus penggunaan utama. Pertama, user dapat memasukkan data secara langsung ke dalam sistem melalui form input yang disediakan, memungkinkan data baru dimasukkan secara manual. Kedua, user dapat mengunggah data menggunakan file, seperti CSV atau Excel, yang mempermudah memasukkan data dalam jumlah besar atau data yang telah dipersiapkan sebelumnya. Ketiga, user dapat mengambil dan melihat hasil input, yaitu melihat, mengunduh, atau mengelola data yang sudah ada di dalam sistem. Interface web dapat dilihat pada Gambar 6 berikut.



**Gambar 6.** User Interface Website

## 4. KESIMPULAN

Kinerja model peramalan energi photovoltaic menggunakan algoritma random forest classification menunjukkan hasil yang baik dengan akurasi yang tinggi dengan nilai 0,96% dan error sebesar 0,04% dalam memprediksi produksi energi PLTS. Hasil peramalan produksi energi PLTS di masa depan menunjukkan potensi yang signifikan dalam mendukung kebutuhan energi terbarukan, dengan estimasi produksi yang konsisten dan dapat diandalkan. Data-data yang mempengaruhi hasil peramalan mencakup berbagai variabel seperti intensitas radiasi sinar matahari, suhu, dan kelembaban yang semuanya memiliki peran penting dalam memastikan akurasi dan reliabilitas model peramalan. Secara keseluruhan, penelitian ini memberikan kontribusi yang berharga dalam pengembangan teknologi peramalan energi terbarukan, khususnya dalam aplikasi PLTS. Selain itu, penggunaan algoritma random forest terbukti efektif dalam mengolah data yang kompleks dan variatif, memberikan hasil yang lebih baik. Implementasi model ini diharapkan dapat mendorong adopsi yang lebih luas dari sumber energi bersih di berbagai sektor. Penelitian ini juga dapat menjadi acuan bagi pengembangan teknologi serupa di masa depan, serta menjadi dasar untuk penelitian lanjutan dalam upaya meningkatkan efisiensi dan efektivitas produksi energi terbarukan.

## REFERENCES

- [1] H. B. Nurjaman and T. Purnama, “Pembangkit Listrik Tenaga Surya (PLTS) Sebagai Solusi Energi Terbarukan Rumah Tangga,” 2022. [Online]. Available: <https://journal.uny.ac.id/index.php/jee>



- [2] J. Homepage, A. Roihan, P. Abas Sunarya, and A. S. Rafika, "Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper," 2019.
- [3] T. P. Carvalho, F. A. A. M. N. Soares, R. Vita, R. da P. Francisco, J. P. Basto, and S. G. S. Alcalá, "A Systematic Siterature Review of Machine Learning Methods Applied to Predictive Maintenance," *Comput Ind Eng*, vol. 137, Nov. 2019, doi: 10.1016/j.cie.2019.106024.
- [4] F. Baharuddin and A. Tjahyanto, "Peningkatan Performa Klasifikasi Machine Learning Melalui Perbandingan Metode Machine Learning dan Peningkatan Dataset," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 11, no. 1, pp. 25–31, Mar. 2022, doi: 10.32736/sisfokom.v1i1.1337.
- [5] H. Rathore, H. K. Meena, and P. Jain, "Prediction of EV Energy consumption Using Random Forest And XGBoost," in Proceedings - 2nd International Conference on Power Electronics and Energy, ICPEE 2023, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICPEE54198.2023.10060798.
- [6] Ardianto, A. Raharjo, and D. Purwitasari, "Random Forest Regression Untuk Prediksi Produksi Daya Pembangkit Listrik Tenaga Surya," *BRILIANT: Jurnal Riset dan Konseptual*, vol. 7, pp. 1058–1075, 2022, doi: <http://dx.doi.org/10.28926;briliant.v7i4.1036>.
- [7] H. Gufron, D. Rusirawan, and L. Widywati, "Forecasting Produksi Energi PLTS 1 kWp Menggunakan Mesin Pembelajaran Dengan Algoritma Support Vector Machine," *Jurnal Tekno Insentif*, vol. 16, no. 2, pp. 79–91, Dec. 2022, doi: 10.36787/jti.v16i2.843.
- [8] M. M. Gokce and E. Duman, "Performance Comparison of Simple Regression, Random Forest and XGBoost Algorithms for Forecasting Electricity Demand," in 3rd International Informatics and Software Engineering Conference, IISEC 2022, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/IISEC56263.2022.9998213.
- [9] D. Liu and K. Sun, "Random forest solar power forecast based on classification optimization," *Energy*, vol. 187, Nov. 2019, doi: 10.1016/j.energy.2019.115940.
- [10] Y. A. Singgalen, "Penerapan Metode CRISP-DM dalam Klasifikasi Data Ulasan Pengunjung Destinasi Danau Toba Menggunakan Algoritma Naïve Bayes Classifier (NBC) dan Decision Tree (DT)," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, pp. 1551–1526, 2023, doi: 10.30865/mib.v7i3.6461.
- [11] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 526–534. doi: 10.1016/j.procs.2021.01.199.
- [12] S. Hani, G. Santoso, S. Subandi, and N. Arifin, "Perencanaan Pembangkit Listrik Tenaga Surya (PLTS) On-Grid Dengan Sistem DC Coupling Berkapasitas 17 kWp Pada Gedung," *Prosiding Seminar Nasional Teknoka*, vol. 5, pp. 156–163, Dec. 2020, doi: 10.22236/teknoka.v5i.300.
- [13] A. Santosa, S. Hani, and G. Santoso, "Perancangan Sistem PLTS Off-grid Kapasitas 100 Wp Sebagai Sumber Energi Alternatif Charging 220 V di Daerah Terdampak Bencana Semeru," 2022.
- [14] N. Aini, A. Lestari, M. N. Hayati, F. Deny, and T. Amijaya, "Analisis Cluster Pada Data Kategorik dan Numerik dengan Pendekatan Cluster Ensemble (Studi Kasus: Puskesmas di Provinsi Kalimantan Timur Kondisi Desember 2017)," *Jurnal EKSPONENTIAL*, vol. 11, no. 2, 2020.
- [15] B. Julian, Muliadi, and Syukri, "Analisis Pengaruh Radiasi Matahari Dan Temperatur Terhadap Daya Keluaran Fotovoltaik Menggunakan SPSS," *Achel Journal of Electrical Engineering and Technology*, vol. 3, 2023.
- [16] H. Johan, N. Utomo, and W. W. Rendy, "Pengaruh Temperatur Udara, Kelembaban Udara, Kecepatan Udara dan Intensitas Cahaya Terhadap Daya Listrik Panel Surya," 2022.
- [17] A. Ferdita Nugraha, Y. Pristyanto, and I. Pratama, "Penanganan Missing Values Untuk Meningkatkan Kinerja Model Machine Learning Pada Data Telemarketing," 2020. [Online]. Available: [www.ejournal.unib.ac.id/index.php/pseudocode](http://www.ejournal.unib.ac.id/index.php/pseudocode)
- [18] E. Blessing and A. K. Hubert, "Normalization and Standardization: Methods to preprocess data to have consistent scales and distributions," 2023. [Online]. Available: <https://www.researchgate.net/publication/377123133>
- [19] W. Aprizal, I. Ariawan, A. Armelita, and N. Tufailah, "Data Scaling Performance on Various Machine Learning Algorithms to Identify Abalone Sex," *Jurnal Teknologi dan Sistem Komputer*, vol. 10, no. 1, pp. 26–31, 2022, doi: 10.14710/jtsiskom.2022.14105.
- [20] R. Bayindir, M. Yesilbudak, M. Colak, and N. Genc, "A novel application of naive bayes classifier in photovoltaic energy prediction," in Proceedings - 16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017, Institute of Electrical and Electronics Engineers Inc., 2017, pp. 523–527. doi: 10.1109/ICMLA.2017.0-108.
- [21] R. Supriyadi, W. Gata, N. Maulidah, and A. Fauzi, "Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah," vol. 13, no. 2, pp. 67–75, Dec. 2020, [Online]. Available: <http://journal.stekom.ac.id/index.php/E-Bisnispage67>
- [22] irwansyah saputra and kristiyanti dinar ajeng, *Machine Learning Untuk Pemula*. bandung: informatika bandung, 2022.
- [23] G. Ashfania, achmad Widodo, T. Warsokusumo, and T. Prahasto, "Penggunaan Algoritma Random Forest untuk Klasifikasi berbasis Kinerja Efisiensi Energi pada Sistem Pembangkit Daya," jakarta, Jul. 2023.
- [24] D. Suryani, A. Labellapansa, M. L.Shiddiqie, and A. Hidayat, "Kajian Klasifikasi Data Mining IQ Siswa SMA Berdasarkan Hasil Intelligence Structure Test Dengan Menggunakan Metode Naive Bayes," *Seminar Nasional Informatika Medis (SNIMed)*, 2019.
- [25] A. Fahrizal, D. Rusirawan, and L. Lidyawati, "Pemodelan Produksi Energi Pembangkit Listrik Tenaga Surya 1000 Wp Dengan Algoritma Naive Bayes," *Jurnal Tekno Insentif*, vol. 16, no. 2, pp. 105–118, Dec. 2022, doi: 10.36787/jti.v16i2.864.
- [26] K. Immanuel Rui Costa, "Pengembangan dan Pembuatan Website: Sebuah Tinjauan Literatur," *palangkaraya*, Apr. 2022. Accessed: Jul. 07, 2024. [Online]. Available: <https://www.researchgate.net/publication/359815377>