

Prediksi Potensi Suatu Wilayah Untuk Menjadi PLTS Dengan Machine Learning

Ratri Ismayanti¹, Wiga Maulana Baihaqi²

¹Universitas Amikom Purwokerto Program Sistem Informasi, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Jawa Tengah, Indonesia

²Universitas Amikom Purwokerto Program Teknologi Informasi, Fakultas Ilmu Komputer, Universitas Amikom Purwokerto, Jawa Tengah, Indonesia

surel: [1ratriismayanti@gmail.com](mailto:ratriismayanti@gmail.com), [2wiga@amikompurwokerto.ac.id](mailto:wiga@amikompurwokerto.ac.id)

Info Artikel

Sejarah artikel:

Diterima 02 Februari 2024

Revisi 29 Juni 2024

Diterima 08 Agustus 2024

Kata kunci:

Liniear Regression, Lasso Regression, Ridge Regression, dan Support Vector Regression

Abstrak

Listrik sudah menjadi kebutuhan pokok sebagian umat manusia karena semua aktivitas hampir berhubungan dengan listrik. Indonesia memiliki beberapa proyek Pembangkit Listrik dan Pembangkit Listrik yang terbesar dihasilkan dari PLTU yang mana bisa dampak yang kita rasakan adalah emisi gas rumah kaca dan polusi udara yang buruk dan juga sangat bergantung pada batu bara sedangkan SDA tersebut tidak bisa diperbarui dengan fakta tersebut jika kita mengurangi penggunaan batu bara maka akan menjadi boomerang bagi Indonesia sendiri. Penelitian ini bertujuan untuk memprediksi wilayah yang dapat berpotensi menjadi Pemangkit Listrik Tenaga Surya dengan pendekatan model regresi *machine learning*. Sehingga diharapkan penelitian ini bisa menjadi acuan dalam pengembangan Pembangkit Listrik Tenaga Surya di Indonesia. Metode yang digunakan adalah *Liniear Regression*(LR), *Lasso Regression*(LR), *Ridge Regression*(RR), dan *Support Vector Regression* (SVR). Koefisien R² untuk radiasi sinar matahari berturut-turut adalah 0.924; 0.910; 0.917; 0.949; dan 0.987.

Penulis yang sesuai:

Ratri Ismayanti
Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Amikom Purwokerto
Email: ratriismayanti@gmail.com

1. PENDAHULUAN

Era yang semakin sadar akan lingkungan, peranan energi terbarukan seperti Pembangkit Listrik Tenaga Surya (PLTS) menjadi semakin krusial. PLTS dianggap sebagai solusi terbaik untuk mengurangi dampak lingkungan dari pembangkit listrik konvensional berbasis bahan bakar fosil. Namun, implementasi yang efisien memerlukan pemahaman mendalam terhadap potensi energi surya di suatu wilayah. Proyek akhir berjudul "ENBAS: Potensi Energi pada Suatu Wilayah dalam Pembangunan PLTS" bertujuan untuk menyelidiki dan mengevaluasi potensi energi surya

yang dapat dimanfaatkan untuk pembangunan PLTS di wilayah tertentu. Analisis ini mencakup faktor-faktor seperti kondisi cuaca, intensitas matahari, dan topografi wilayah.

Pentingnya beralih ke sumber energi terbarukan untuk mengurangi emisi gas rumah kaca dan memitigasi perubahan iklim global menjadi latar belakang utama proyek ini. PLTS menjadi teknologi terdepan dalam portofolio energi terbarukan, tetapi untuk implementasi yang sukses, diperlukan analisis mendalam terhadap potensi energi surya di suatu wilayah. Dengan pemahaman yang baik, dapat dirancang sistem PLTS yang sesuai, memaksimalkan penggunaan energi surya dan mengoptimalkan keberlanjutan proyek energi terbarukan. Proyek ini diharapkan memberikan pandangan mendalam dan solusi praktis untuk membangun PLTS yang efisien dan berkelanjutan di wilayah yang diteliti, serta berkontribusi pada pengembangan infrastruktur energi terbarukan.

Solusi yang diusulkan, ENBAS: Potensi Energi pada Suatu Wilayah dalam Pembangunan PLTS, melibatkan penggunaan kecerdasan buatan (AI) untuk menghitung dan visualisasi potensi energi surya. Dengan menggabungkan data topografi, cuaca, dan faktor lainnya, dashboard AI ini memberikan informasi tentang lokasi terbaik untuk pembangunan PLTS. Solusi ini diharapkan dapat menggantikan proses manual, membantu meningkatkan efisiensi perencanaan proyek PLTS, dan mendukung pencapaian tujuan energi terbarukan. Keberhasilan solusi ini akan memberikan kontribusi positif dalam menghadapi tantangan global terkait energi dan lingkungan.

Model pembelajaran mesin (ML) dapat membentuk korelasi non-linier antara variable dependen dan variable independen. Penelitian pembuatan model Forecasting produksi energi pada PLTS 1 kWp ITENAS yang berdasarkan data produksi energi harian secara deret waktu, untuk memprediksi periode selanjutnya, Gufron dkk (2022). Algoritma Support Vector Machine (SVM) digunakan dalam pembelajaran mesin untuk membentuk model yang dapat memprediksi masa depan. Dengan pengembangan SVM, analisis regresi berubah menjadi Support Vector Regression (SVR), sebuah metode pembelajaran yang berkembang dari teori statistik, menonjol karena kemampuannya yang superior dalam mengklasifikasikan data non-linier dibandingkan dengan metode lainnya, Shi dkk dalam Gufron dkk (2022).

Banyak faktor yang memengaruhi produksi energi surya, termasuk radiasi surya, tutupan awan, suhu, kelembaban, tekanan atmosfer, dan kecepatan angin, Wang dkk (2020). Model machine learning yang berbasis pada algoritma regresi random forest (RF) dikembangkan untuk memprediksi Efisiensi Energi Konversi (PCE) dengan menggunakan deskriptor elektronik seperti tingkat energi orbital molekuler dan celah pita dari donor dan akseptor, lee dkk dalam Shutar (2023). Penelitian ini menggunakan empat algoritma untuk memperoleh hasil prediksi wilayah berpotensi menjadi PLTS sehingga mendekati akurat. Berikut ini dijeaskan empat algoritma yang digunakan dalam penelitian ini; (1) Regresi Linier, Regresi linier digunakan untuk memodelkan hubungan linier antara berbagai faktor (seperti lokasi geografis, cuaca, dan topografi) dengan potensi energi matahari di wilayah tersebut.

Dengan menggunakan data historis atau pengamatan, dapat membangun model regresi linier untuk memprediksi potensi energi surya di wilayah tersebut. Analisis regresi dilakukan dengan menganalisis keterkaitan antara dua atau lebih variabel independen, yang juga dikenal sebagai variabel bebas atau predictor, Luthfiarta dkk (2020). Rekomendasi terkait pengembangan penelitian prediksi melibatkan penggunaan berbagai atribut yang relevan seperti suhu, kelembapan udara, tekanan udara, dan kecepatan angin. Diharapkan pendekatan ini dapat menghasilkan pola yang lebih rinci dan prediksi yang lebih akurat, Luthfiarta dkk(2020). (2) *Lasso Regression* dan *Ridge Regression* dapat berguna dalam konteks proyek ini untuk menangani masalah *overfitting*. Mungkin ada banyak faktor yang dapat mempengaruhi potensi energi, dan regresi regulerisasi dapat membantu memilih variabel yang paling relevan atau mengatasi masalah multikolinearitas.

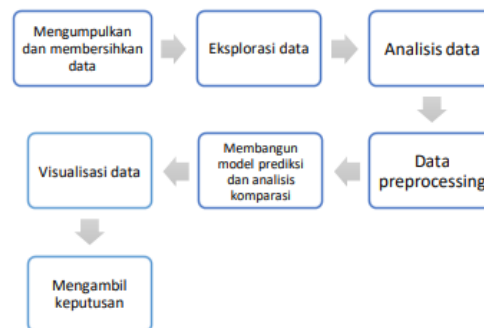
Regresi Ridge digunakan untuk mengatasi *overfitting* dan mengontrol variasi model dengan memasukkan istilah hukuman berupa jumlah kuadrat koefisien ke dalam fungsi tujuan yang diestimasi, Wijayanti (2023). Lasso menggunakan hukuman berupa jumlah mutlak koefisien pada fungsi tujuan yang diestimasi, Wijayanti (2023). (3) *Support Vector Regression*, SVR dapat digunakan untuk membangun model regresi yang lebih kompleks dan dapat menangani pola nonlinier dalam data. Jika hubungan antara variabel dan potensi energi surya tidak linier, SVR bisa menjadi pilihan yang baik. SVR diteliti sebagai alternatif untuk mendekati analisis teknik yang rumit, Aeni (2020). Metode SVM dapat diterapkan pada situasi regresi dan data non-linier yang melibatkan jumlah data input yang besar, Caraka (2017). (4) Decision Tree dapat membantu mengidentifikasi pola dan interaksi kompleks antara berbagai faktor yang mempengaruhi potensi energi surya sehingga dapat membuat pohon keputusan untuk memahami faktor-

faktor mana yang paling berpengaruh. *Decision Tree* adalah suatu teknik yang sangat efektif dan terkenal dalam melakukan klasifikasi dan prediksi, Muzakir (2016).

2. METODE

Penelitian ini dilaksanakan dengan tahapan berikut: (1) Pengumpulan Data: Dataset diperoleh dari sumber NASA untuk memastikan integritas dan kualitas data. Informasi ini sangat penting untuk mendukung analisis yang akurat dan valid. (2) Pemrosesan Data: Proses preprocessing melibatkan normalisasi data untuk memastikan konsistensi dan meminimalkan efek outlier. Teknik-teknik seperti *Linear Regression*, *Lasso Regression*, dan *Ridge Regression* digunakan untuk mengatasi masalah multicollinearity dan memperbaiki model. (3) Penanganan Nilai yang Hilang: Metode imputasi data yang canggih, seperti *Support Vector Regression (SVR)*, digunakan untuk menangani nilai yang hilang secara efisien dan mengoptimalkan penggunaan data yang tersedia. (4) Pembagian Data: Dataset dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian, untuk melakukan validasi silang dan menghindari overfitting model. (5) Pengembangan Model: Proses pengembangan model melibatkan implementasi *Linear Regression* untuk mengevaluasi hubungan linear antara variabel, *Lasso* dan *Ridge* untuk melakukan regularisasi, *SVR* untuk menangani nonlinearitas, dan *Decision Tree Regression* untuk menangani struktur data yang kompleks. (6) Validasi dan Evaluasi: Performa model dievaluasi menggunakan sejumlah metrik, termasuk *Mean Squared Error (MSE)*, *Mean Absolute Error (MAE)*, dan *Root Mean Squared Error (RMSE)*, untuk mengukur sejauh mana model mampu memprediksi radiasi sinar matahari.

Secara keseluruhan, langkah-langkah untuk memperoleh hasil prediksi dari model machine learning dalam penelitian ini diberikan melalui suatu diagram alir yang tergambar di Gambar 1.



Gambar 1. Diagram Alir Penelitian

1.1 Dataset

Dataset diperoleh melalui website resmi NASA yang dimana didalam dataset tersebut sudah terdapat factor data yang dibutuhkan seperti; suhu udara, kelembapan permukaan, intensitas radiasi, fluks gelombang pendek masuk permukaan, total ozon kolom, total uap air yang dapat jatuh sebagai presipitasi, dan kecepatan angin.

1.2 Model Prediksi

Linear Regression adalah metode statistik yang digunakan untuk menemukan hubungan linier antara variabel dependen (output) dan satu atau lebih variabel independen (input).

$$y = b_0 + b_1x_1 + b_2 x_2 + \dots + b_nx_n,$$

di mana y adalah variabel dependen, x_1, x_2, \dots, x_n adalah variabel independen, dan $b_0, b_1, b_2, \dots, b_n$ adalah koefisien regresi.

Lasso Regression adalah variasi dari regresi linear dengan penambahan regularisasi L1, yang membantu mencegah overfitting dan dapat menghasilkan model yang lebih sederhana.

$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n + \lambda \sum_{i=1}^n b_i^2$$

dimana λ adalah parameter penalti.

Support Vector Regression adalah metode regresi yang menggunakan *Support Vector Machines* (SVM) untuk memprediksi nilai kontinu. Tujuannya adalah menemukan hyperplane yang memiliki deviasi maksimum dari data target. *Decision Tree* menggunakan struktur pohon keputusan untuk memodelkan hubungan antara input dan output. Pohon dibagi berdasarkan kondisi pada fitur-fitur input.

Dalam evaluasi performa model, dua metrik umum yang digunakan adalah *Mean Squared Error* (MSE) dan *Mean Absolute Error* (MAE). Untuk mengukur ketepatan prediksi, MSE dan MAE dihitung sebagai berikut:

$$\text{Mean Squared Error (MSE)} : \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (1)$$

Dimana y_i adalah nilai actual, \hat{y}_i adalah nilai prediksi, dan n adalah jumlah observasi.

$$\text{Mean Absolute Error (MAE)} : \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

Dimana y_i adalah nilai actual, \hat{y}_i adalah nilai prediksi, dan n adalah jumlah observasi.

3. HASIL DAN PEMBAHASAN

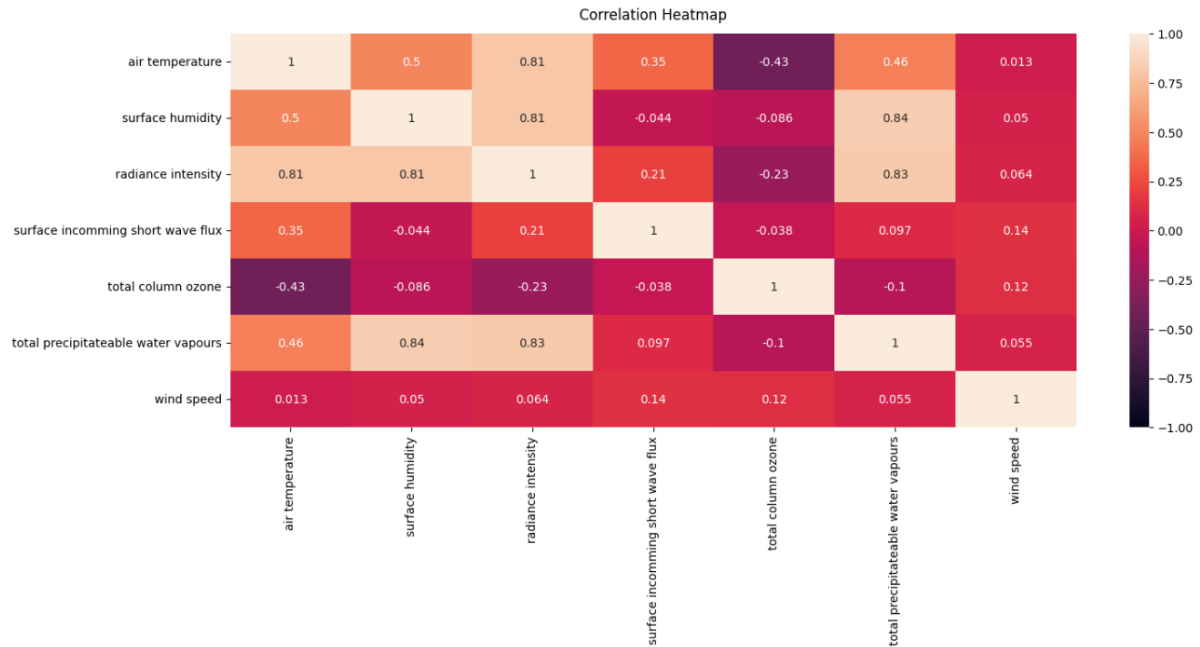
Berdasarkan **Gambar 2** data tersebut adalah dataset yang dibutuhkan untuk melakukan pemodelan dan menentukan algoritma yang tepat dan sesuai untuk pembuatan machine learning.

	time and date	air temperature	surface humidity	radiance intensity	surface incoming short wave flux	total column ozone	total precipitates	water vapours	wind speed
0	8/1/2019 0:00	20.733210	0.015783	336.418793	2.65332	269.776154	19.257727	6.649365	
1	8/1/2019 1:00	20.610926	0.015502	333.328033	107.12500	268.784729	18.675245	6.926465	
2	8/1/2019 2:00	20.968317	0.015116	330.606750	320.00000	268.444275	18.060242	7.402035	
3	8/1/2019 3:00	23.868128	0.012997	335.808411	545.50000	269.155518	17.486549	7.386476	
4	8/1/2019 4:00	26.353357	0.010358	343.027130	747.00000	270.530609	17.188797	7.045732	
...	
13147	1/29/2021 19:00	1.927454	0.001664	196.981537	0.00000	251.769577	2.121896	3.679206	
13148	1/29/2021 20:00	2.280206	0.001548	200.435287	0.00000	251.132355	2.380484	3.625995	
13149	1/29/2021 21:00	2.690790	0.001436	203.228104	0.00000	250.384811	2.547078	3.631343	
13150	1/29/2021 22:00	3.166894	0.001345	204.969635	0.00000	249.664856	2.747611	3.699098	
13151	1/29/2021 23:00	3.624536	0.001290	207.033783	0.00000	248.726105	3.057377	3.913473	

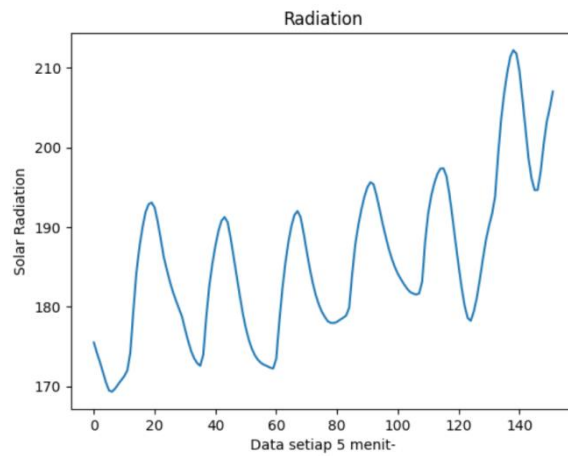
13152 rows x 8 columns

Gambar 2. Dataset

Korelasi yang terjadi antara dataset intensitas radiasi dengan suhu udara diinterpretasikan dalam grafik *Heat map* yang ditampilkan pada **Gambar 3**. Nilai koefisien korelasi antara intensitas radiasi dengan suhu udara, kelembapan permukaan, fluks gelombang pendek masuk permukaan, total kolom ozon, total uap air yang dapat diendapkan, dan kecepatan angin adalah 0,5 ; 0,81 ; ----0,044 ; -0,086 ; 0,84 ; 0,05.



Gambar 3. Heat Map Koefisien Korelasi intensitas radiasi dengan perubahan cuaca

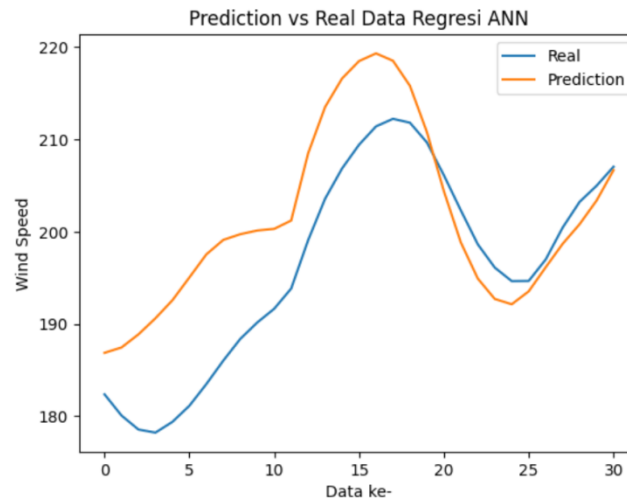


Gambar 4. Grafik Solar Radiasi dengan Data setiap 5 menit

Tabel 1. Perhitungan Evaluasi Model Prediksi Radiasi Sinar Matahari

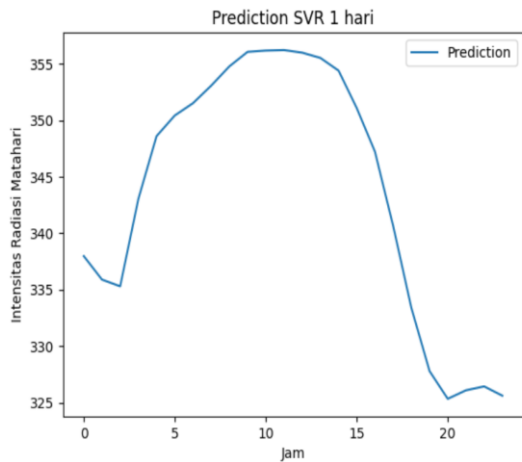
Model	RMSE Data Training	RMSE Data Testing	R ² Score
Linier	12.62336974801681	16.22972648106027	0.924
Lasso	13.754860653128372	17.649528642943594	0.910
Ridge	13.198284229649001	17.170457815302107	0.917
SVR	10.608602166809204	12.31692185637924	0.949
DTR	0.0	12.13256143166338	0.987

Hasil algoritma yang digunakan dalam penelitian ini sebagai berikut:

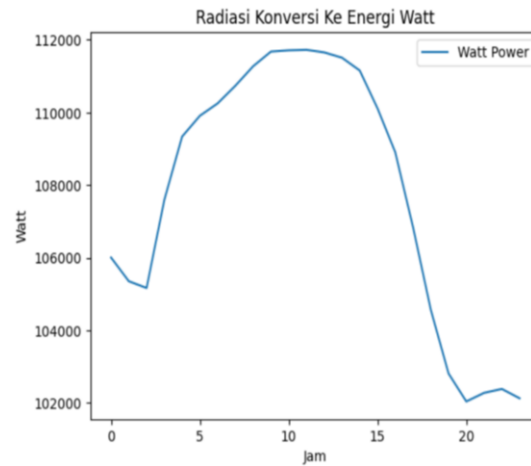
**Gambar 5.** Model Prediksi Terhadap Real Data Regresi ANN

Dalam **Gambar 3**, terlihat bahwa koefisien korelasi antara dataset intensitas radiasi dan dataset perubahan cuaca paling tinggi ditemukan saling berkaitan. Setelah menjalankan percobaan dengan menggunakan data uji pada model *Linier Regression*, *Lasso Regression*, *Ridge Regression*, *Support Vector Regression*, dan *Decision Tree* nilai yang didapatkan adalah RMSE dan R² dapat dilihat pada **Tabel 1**. Model linier ini memiliki tingkat kesalahan yang relatif rendah pada data training dan testing. R² Score yang tinggi menunjukkan bahwa model ini cukup baik dalam menjelaskan variasi dalam data. Model Lasso memiliki tingkat kesalahan yang sedikit lebih tinggi daripada model linier pada kedua data training dan testing. Meskipun demikian, R² Score yang tinggi menunjukkan kemampuan model untuk memberikan penjelasan yang baik terhadap variasi dalam data. Model Ridge memberikan hasil yang serupa dengan model Lasso, dengan tingkat kesalahan yang sedikit lebih tinggi pada data testing. R² Score yang tinggi menunjukkan kemampuan model untuk menjelaskan variasi dalam data. Model SVR memberikan tingkat kesalahan yang lebih rendah pada kedua data training dan testing dibandingkan dengan model linier, Lasso, dan Ridge. R² Score yang tinggi menunjukkan bahwa model ini sangat baik dalam menjelaskan variasi dalam data. Model DTR pada data ini kemungkinan besar mengalami overfitting, karena RMSE pada data training adalah 0.0, yang berarti model hampir sempurna dalam memprediksi data training. Sebagai solusi, dapat mencoba melakukan pemangkasan pohon (pruning)

dengan mengatur parameter seperti kedalaman maksimum pohon, jumlah minimum sampel untuk membagi node, dan jumlah minimum sampel di setiap daun. Tujuannya adalah untuk mengendalikan kompleksitas model dan mencegah overfitting. Visualisasi model pada data testing prediksi ditunjukkan pada **Gambar 6** dan **Gambar 7**.



Gambar 6. Hasil Prediksi dari Data Testing



Gambar 7. Visualisasi Watt per-Jam

Pada **Gambar 6** dan **Gambar 7** dapat dilihat bahwa model yang digunakan membuat data testing terprediksi dengan baik sehingga bisa dikatakan bahwa model yang dirancang berhasil.

4. KESIMPULAN

Dari penelitian ini semoga dapat memberikan inspirasi bagi peneliti lain agar bisa menjadi alternatif yaitu bisa mengurangi menggunakan bahan fosil yang tidak dapat didaur ulang sehingga dapat membantu pembuatan PLTS dan bisa menggunakan prediksi machine learning yang mana akurasi mencapai 0.987.

REFERENSI

- [1] Caraka, R. E., Yasin, H., & Basyiruddin, A. W. (2017). Peramalan Crude Palm Oil (CPO) Menggunakan Support Vector Regression Kernel Radial Basis. *Jurnal Matematika*, 7(1), 43. <https://doi.org/10.24843/jmat.2017.v07.i01.p81>
- [2] Gufron, H., Rusirawan, D., & Widyawati, L. (2022). Forecasting Produksi Energi PLTS 1 kWp Menggunakan Mesin Pembelajaran Dengan Algoritma Support Vector Machine. *Jurnal Tekno Insentif*, 16(2), 79–91. <https://doi.org/10.36787/jti.v16i2.843>
- [3] Luthfiarta, A., Febriyanto, A., Lestiawan, H., & Wicaksono, W. (2020). Analisa Prakiraan Cuaca dengan Parameter Suhu, Kelembaban, Tekanan Udara, dan Kecepatan Angin Menggunakan Regresi Linear Berganda. *JOINS (Journal of Information System)*, 5(1), 10–17. <https://doi.org/10.33633/joins.v5i1.2760>
- [4] Muzakir, A., & Wulandari, R. A. (2016). Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree. *Scientific Journal of Informatics*, 3(1), 19–26. <https://doi.org/10.15294/sji.v3i1.4610>
- [5] Nur 'aeni¹, U., Prasasti², A. L., & Kallista³, M. (2020). Prediksi Jumlah Penumpang Dan Penambahan Gerbong Kereta Api Menggunakan Metode Support Vector Regression (Svr) Prediction of Total Passengers and Addition of Railway Train Using Support Vector Regression (Svr) Method. 7(2), 4919–4926.
- [6] Suthar, R., Abhijith, T., Sharma, P., & Karak, S. (2023). Machine learning framework for the analysis and prediction of energy loss for non-fullerene organic solar cells. *Solar Energy*, 250, 119-127. <https://doi.org/10.1016/j.solener.2022.12.029>
- [7] Wang, H., Liu, Y., Zhou, B., Li, C., Cao, G., Voropai, N., & Barakhtenko, E. (2020). Taxonomy research of artificial intelligence for deterministic solar power forecasting. *Energy Conversion and Management*, 214(May), 112909. <https://doi.org/10.1016/j.enconman.2020.112909>
- [8] Wijayanti, A. (2023). Penanganan Masalah Multikonlinearitas dengan Regresi Ridge, Lasso, dan Elastic-net pada Kasus Balta Stunting di Indonesia. *Journal of Engineering Research*.