

PERBANDINGAN MODEL PREDIKSI RADIASI MATAHARI BERBASIS MESIN PEMBELAJARAN PADA STASIUN METEOROLOGI FATMAWATI SOEKARNO BENGKULU

COMPARISON OF SOLAR RADIATION PREDICTION MODELS BASED ON MACHINE LEARNING AT METEOROLOGICAL STATION OF FATMAWATI SOEKARNO BENGKULU

Dodi Ardiansyah¹, Diah Novita Astuti²

¹Stasiun Meteorologi Fatmawati Soekarno Bengkulu
Jl. Depati Payung Negara Kel. Pekan Sabtu Kec. Selebar Bengkulu, 38213
*Email : ardidodi80@gmail.com

²Stasiun Meteorologi Klas I Juanda Sidoarjo
Jl. Raya Bandara Juanda, Pranti, Kec. Sedati, Kabupaten Sidoarjo, Jawa Timur 61253
*Email : novita_ayu@ymail.com

Naskah masuk: 15 Februari 2023 Naskah diperbaiki: 14 Maret 2023 Naskah diterima: 14 April 2023

ABSTRAK

Radiasi matahari adalah sumber energi terbesar dan memiliki peran dalam keseimbangan radiasi permukaan, siklus hidrologi, fotosintesis vegetasi, cuaca dan iklim. Sangat penting untuk menganalisis radiasi matahari dalam berbagai keperluan. Makalah ini bertujuan untuk mempelajari dan mengevaluasi kelayakan metode-metode didalam mesin learning dalam membuat prediksi radiasi matahari. Parameter Tekanan Udara, Temperatur Udara, Dew Point, Kelembaban Udara yang diukur bersama dengan Radiasi Matahari digunakan untuk membuat prediksi. Tiga metode yang di gunakan dalam makalah ini yaitu Linear Regression (LR), Random Forest Regressor (RFR), dan Decision Tree Regressor (DTR). Untuk menentukan kinerja hasil prediksi dilakukan evaluasi dengan tiga matrik statistic yaitu Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) dan R-Squared (R^2). Hasil Prediksi menunjukkan bahwa metode Random Forest Regressor memberikan nilai prediksi yang paling baik dari dua metode lainnya yang dicoba dengan nilai MAE 51,4 nilai RMSE 106,0 dan R^2 0,87.

Kata Kunci : Linear Regression, Random Forest Regressor, Decision Tree Regressor, Root Mean Squared Error, R-Squared

ABSTRACT

Solar radiation is the biggest energy source of the earth and has a role in the balance of surface radiation, the hydrological cycle, photosynthesis of vegetation, weather and climate. Crucial to analyze of solar radiation in various purposes. This paper aims to study and evaluate a appropriateness of methods in a machine learning for predictions of solar radiation. Parameters of Air Pressure, Air Temperature, Dew Point, Air Humidity which measured along with Solar Radiation are used to make predictions. The three methods used in this paper are Linear Regression (LR), Random Forest Regressor (RFR), and Decision Tree Regressor (DTR). To determine the performance of the prediction results, three statistical matrices were evaluated, it is Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R-Squared (R^2). Prediction results show that the Random Forest Regressor method provides the best predictive value of the other two methods have tried with an MAE value of 51.4, RMSE 106.0 and R^2 0.87.

Keywords : Linear Regression, Random Forest Regressor, Decision Tree Regressor, Root Mean Squared Error, R-Squared

1. Pendahuluan

Matahari mempunyai peranan penting dalam kehidupan di bumi yang dapat dimanfaatkan sebagai sumber energi terbarukan. Energi matahari dihantarkan ke ruang angkasa dalam bentuk radiasi elektromagnetik yang memiliki kecepatan sangat tinggi yaitu 3×10^8 m/s dan dapat melewati ruang hampa. Dalam peranannya sebagai solusi energi terbarukan yang layak dalam menggantikan sumber daya saat ini terutama sumber daya listrik, keberadaan data radiasi matahari sangat signifikan digunakan diberbagai sektor seperti dalam konversi dan pembangkitan energi dari sinar matahari, pemanas air, penyulingan air dan meteorologi [1]. Radiasi Matahari sendiri merupakan pancaran energi yang berasal dari proses thermonuklir yang terjadi di matahari. Intensitas radiasi Matahari yang diterima Bumi (intensitas insolasi) bergantung pada lintang atau letak suatu tempat di permukaan bumi [2].

Insolasi biasanya dinyatakan dalam satuan Watt/m²-detik yang mengandung makna intensitas atau kekuatan. Indonesia mempunyai sumber energi surya yang berlimpah dengan intensitas radiasi matahari rata-rata sekitar 4.8 kWh/m² per hari hal ini di karenakan Indonesia terletak pada garis khatulistiwa [3]. Memahami perilaku radiasi matahari di suatu wilayah sangatlah penting sebagai optimalisasi penggunaan energi matahari. Prediksi radiasi matahari yang akurat sangatlah penting dalam perspektif operasional. Dalam beberapa model prediksi, karakteristik unsur meteorologi dijadikan acuan sebagai estimasi luminositas matahari dan radiasi matahari, dimana suhu udara mempunyai korelasi tertinggi [4].

Banyak studi dalam beberapa literatur yang mengeksplor prediksi data radiasi matahari menggunakan beberapa tehnik seperti prediksi radiasi matahari setiap jam dengan algoritma mesin learning menggunakan metode Multilayer Feed-Forward Neural Network (MFFNN), *K*-Nearest Neighbors (*K*-NN) dengan hasil model estimasi paling sukses yang dikembangkan adalah MFFNN [5]. Penggunaan teknik Artificial Neural Network lainnya dalam pemodelan dan prediksi radiasi matahari dilakukan untuk 4 kota di India yang mana hasilnya menunjukkan prediksi menggunakan jaringan syaraf tiruan lebih baik daripada model konvensional[6]. Penggunaan jaringan saraf tiruan yang mampu menafsirkan prediksi radiasi matahari yang efisien dengan akurasi kinerja yang baik menggunakan dua indikator statistik Mean Absolute Error (MAE) dan Mean Squared Error (MSE)[7]. Pengembangan sistem inferensi

neuro-fuzzy adaptif (ANFIS) menggunakan rata-rata suhu minimum bulanan, rata-rata suhu maksimum bulanan, durasi sinar matahari dan data meteorologi lainnya dinyatakan sangat efektif dan baik untuk estimasi radiasi matahari [8]. Penggunaan model machine learning diterapkan untuk menemukan hubungan dalam pengenalan pola dan masalah klasifikasi dimana tidak ada representasi antara input dan output, dalam masalah data mining dan prediksi [9]. Alasan lebih lanjut mengapa machine learning menjadi populer adalah bahwa mereka dapat merumuskan nonlinier data secara numerik, hanya berdasarkan data historis tanpa memerlukan pengetahuan tentang proses fisik yang mendasarinya [10]. Pada model machine learning terdapat metode supervised learning seperti pengklasifikasian, regresi linier, decision tree, random forest regressor [11, 12, 13].

Tujuan penelitian ini adalah mengevaluasi prediksi radiasi matahari menggunakan pemodelan machine learning dengan metode linear regression, random forest regressor dan decision tree dengan menggunakan sampel data dari peralatan AWOS (Automated Weather Observing System) dua bulan dari time series harian yang kemudian diinterpretasikan secara matematis sebagai ekspektasi bersyarat untuk memilih model yang tepat.

2. Metode Penelitian

Data

Pada penelitian ini, data sampel yang digunakan adalah data time series per menit selama 2 bulan yaitu data temperature udara, dew point, kelembaban udara, tekanan udara, radiasi matahari. Parameter respon yang akan diprediksi adalah radiasi matahari. Data sampel tersebut di dapatkan dari hasil rekaman peralatan AWOS (Automatic Weather Observing System) pada Stasiun Meterologi Fatmawati Soekarno Bengkulu) selama 2 bulan yaitu Pebruari dan Maret 2018.

Metode

Regresi Linear

Regresi linear adalah analisa untuk mempelajari dan mengukur hubungan yang terjadi antara dua variabel atau lebih [14]. Regresi linear merupakan salah satu tehnik yang paling penting dan banyak di gunakan. Hal tersebut dikarenakan model regresi linear adalah model yang paling sederhana dan mudah untuk menginterpretasikan hasil. Penelitian ini menggunakan multiple linear regresi dengan persamaan multiple regresi :

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p$$

Di mana :

Y = variable tak bebas

b_0 = Konstanta

$b_1 = b_2 = \dots = b_p$ = Koefisien regresi

$X_1 = X_2 = \dots = X_p$ = Variabel bebas

Di mana variabel prediksi yaitu radiasi matahari sebagai variabel dependen kemudian variable prediktor atau independennya yaitu temperature udara, dew point, tekanan udara dan kelembaban.

Decision Tree dan Random Forest Regressor

Metode *random Forest Regressor* adalah algoritma machine learning populer yang termasuk dalam teknik supervised learning[15]. Metode Random Forest merupakan salah satu metode dalam *Decision Tree*. Algoritma ini dapat digunakan untuk masalah klasifikasi dan regresi. *Decision tree* sendiri yaitu sebuah diagram alir yang berbentuk seperti pohon yang memiliki sebuah root node yang digunakan untuk mengumpulkan data, sebuah inner node yang berada pada root node yang berisi tentang pertanyaan tentang data dan sebuah leaf node yang digunakan untuk memecahkan masalah serta membuat keputusan [16]. Jika nilai respons pada node adalah y_1, \dots, y_n , kriteria pemisahan yang khas adalah rata-rata kuadrat pada node yang dinyatakan sebagai :

$$Q = \frac{1}{n} \sum_{i=1}^n (y_i - y)^2$$

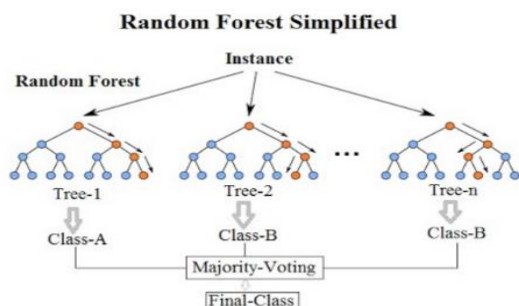
Dimana,

$$y = \frac{1}{n} \sum_{i=1}^n y_i$$

n = Jumlah data

y_i = Nilai respon

y = Merupakan nilai prediksi pada simpul (rata-rata nilai respons).



Data Set

Kualitas data sangatlah penting mengingat bisa saja terjadi kesalahan pada instrument peralatan observasi yang datanya digunakan untuk penelitian. Pembersihan data merupakan proses penting yang akan meningkatkan kualitas data mentah yang akan diteliti . Pada bagian ini kumpulan data yang diambil dan digunakan terlebih dahulu dicari yang mempunyai nilai NULL ataupun error kemudian menghapusnya. Periode waktu data yang digunakan tanggal 1 Pebruari sampai dengan 31 maret 2018. Untuk lebih mudah melihat distribusi data maka divisualisasikan dalam bentuk grafik distribusi dan box plot.

Seleksi Variabel

Pemilihan variable dalam membuat sebuah model sangatlah penting karena variable dapat mencerminkan karakteristik populasi yang akan diteliti. Dari data set yang tersedia diperlukan seleksi variable yang digunakan sebagai variable dependent dan independent. Variabel dependent adalah variabel yang disebabkan/ dipengaruhi oleh adanya variabel bebas/variabel independent dalam hal ini radiasi matahari merupakan variable dependent atau target prediksi. Sedangkan variable independent adalah variable bebas yang mempengaruhi perubahan pada variable dependent atau disebut juga sebagai prediktor, data temperatur udara, dew point, kelembaban udara, tekanan udara digunakan sebagai variable independent. Dari data variabel yang coba diujikan, terlihat korelasi terkuat adalah variable radiasi matahari dan variabel temperature udara sebesar 0.83.

Evaluasi Model

Pertama, membandingkan data radiasi matahari dengan data temperatur udara, dew point, kelembaban udara, tekanan udara. Memplot grafik untuk membantu memahami & memvisualisasikan bagaimana tepatnya kumpulan data yang tersedia. Dalam suatu model tidak ada kriteria khusus yang di gunakan untuk melihat evaluasi kinerja sebuah model. Untuk memaksimalkan kinerja model, meminimalkan fungsi kerugian yang telah ditentukan sebelumnya dan menghasilkan hasil yang lebih baik dengan lebih sedikit kesalahan digunakan tuning hyperparameter. Kinerja dari suatu model dapat di lihat dengan memplot estimasi time series pada real time series [17]. Evaluasi model bertujuan untuk menilai sejauh mana kemampuan penafsiran dapat di percaya. Ada beberapa indikator dan metode yang dapat

digunakan untuk mengevaluasi model secara statistik[16]. Hasil dari model dapat dievaluasi nilai errornya dengan menggunakan matrik statistik yaitu MAE (Mean Absolute Error) dan RMSE (Root Mean Squared Error) yang diperlukan untuk melihat kinerja sebuah model dimana semakin dekat nilainya dengan 0 maka semakin baik hasil prediksinya. Begitu juga dengan R^2 jika nilainya mendekati 1 maka hasil prediksi dan observasi berhubungan erat. MAE (Mean Absolute Error) adalah rata-rata selisih mutlak nilai sebenarnya (aktual) dengan nilai prediksi (peramalan). Dengan rumus :

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - x|$$

n = Jumlah data
 xi = Nilai data aktual
 x = Nilai data prediksi

RMSE (Root Mean Squared Error) adalah jumlah dari kesalahan kuadrat atau selisih antara nilai sebenarnya dengan nilai prediksi yang telah ditentukan. Dengan rumus :

$$RMSE = \sqrt{\frac{(y' - y)^2}{n}}$$

n = Jumlah data
 y' = Nilai data prediksi
 y = Nilai data aktual

R square merupakan suatu nilai yang memperlihatkan seberapa besar variabel independen (eksogen) mempengaruhi variabel dependen (endogen). Dengan rumus :

$$R^2 = 1 - \frac{SS\ Error}{SS\ Total} = 1 - \frac{\sum(y_i - y)^2}{\sum(y_i - \bar{y})^2}$$

SS Error = Nilai variasi dari residu
 SS Total = Nilai variasi total
 yi = Observasi respon ke-i
 y = Rata – rata
 yj = Prediksi respon ke-i

3. Hasil dan Pembahasan

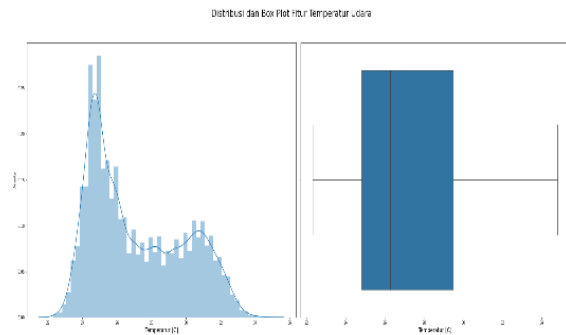
Analisis Data

Nilai maksimum, minimum, mean dan standar deviasi untuk variabel temperatur udara, dew point, tekanan udara, kelembaban, radiasi matahari dapat dilihat pada tabel 1. Untuk data yang digunakan merupakan data permenit.

Parameter	Maks	Min	Mean	Sdv
Temperatur	34.83	22.33	27.13	2.70
Dew Point	27.17	18.11	24.04	0.94
Tek Udara	1015.08	1004.29	1009.2	1.78
Kelembaban	100.00	41.10	84.11	12.38
Radiasi	1263.92	0	197.83	296.12

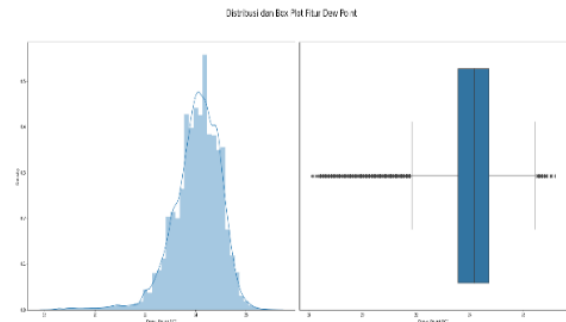
Tabel 1. Deskripsi statistik data

Dari sampel data ditabel terlihat range suhu udara dari 22.44 °C sampai dengan 34.83 °C rata-rata suhu udara 29.12 °C dengan standar deviasi 2.43. Nilai titik embun minimum 18.11 °C sampai 27.17 °C rata-rata titik embun 24.05 dengan standar deviasi 1.06. Tekanan udara mempunyai variasi 1004.29 mb sampai 1015.08 mb dengan rata-rata 1008.9 dan standar deviasinya 1.93. Kelembaban udara mempunyai nilai 41% sampai 100% dengan rata-rata 74.82 dan standar deviasi 11.29. Sedangkan untuk radiasi matahari maksimum mempunyai nilai 1263.92 w/m² nilai terkecil radiasi matahari 1.29 w/m² nilai rata-rata 417.22 dan standar deviasi sebesar 305.61.



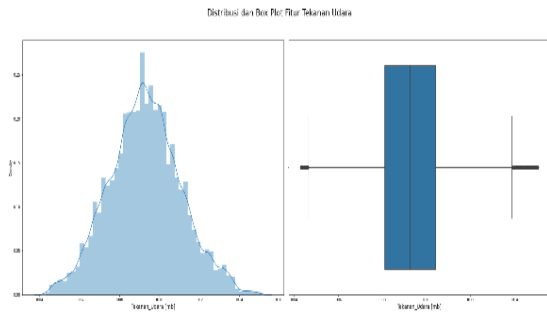
Gambar 1. Density plot dan box plot temperatur udara

Pada gambar terlihat temperatur udara berkisar 22°C sampai 34°C, dilihat dari density temperature udara lebih banyak terjadi pada suhu berkisar 25°C.



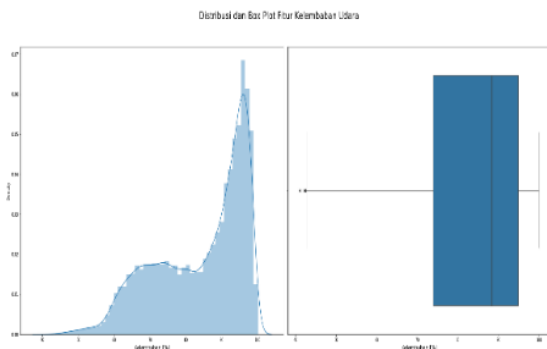
Gambar 2. Density plot dan box plot titik embun

Dari hasil plotting terlihat distribusi titik embun berkisar 18°C hingga 27°C. Suhu titik embun berdasarkan density lebih sering terjadi pada suhu 24°C hingga 25°C.



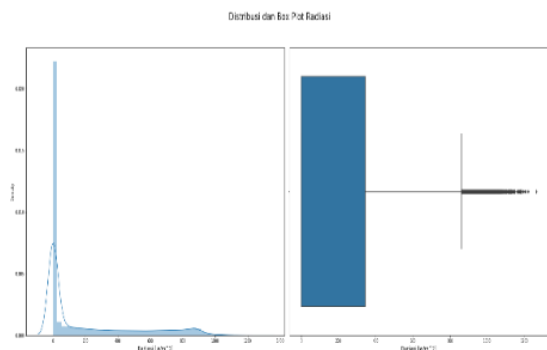
Gambar 3. Density plot dan box plot tekanan udara

Pada grafik terlihat tekanan udara yang terjadi berkisar 1004mb sampai dengan 1015mb. Dilihat dari density tekanan udara lebih sering terjadi berkisar 1009mb.



Gambar 4. Density plot dan box plot kelembaban udara

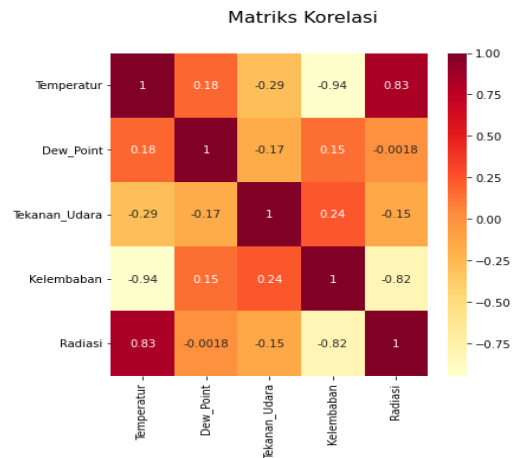
Berdasarkan gambar terlihat kelembaban yang terjadi berkisar 40-100 %. Dari nilai kepadatan terlihat kelembaban yang banyak terjadi berkisar 95% hingga 100%.



Gambar 5. Density plot dan box plot radiasi matahari

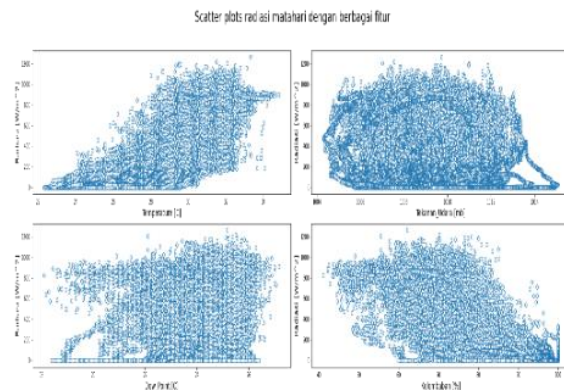
Berdasarkan gambar di atas untuk radiasi matahari terlihat radiasi terjadi mulai dari 0 hingga lebih dari 1000 w/m². Dari nilai

kepadatan radiasi terbanyak 0 w/m² hal ini dikarenakan pada malam hari tidak adanya radiasi matahari.



Gambar 6. Matriks korelasi fitur-fitur data

Matriks korelasi dari berbagai fitur menunjukkan nilai korelasi yang mempunyai hubungan lebih erat yaitu radiasi matahari dan fitur temperature udara. Hal tersebut di tunjukkan dengan besarnya nilai korelasi 0.83 yang mengindikasikan terdapat korelasi yang dekat antara radiasi dan temperatur udara.



Gambar 7. Scatter plot radiasi matahari dengan berbagai fitur

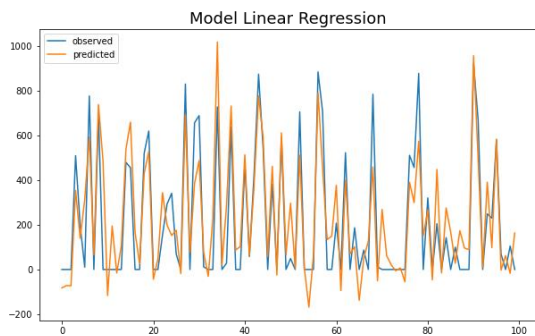
Dari gambar scatter plot terlihat hanya fitur temperatur udara yang mempunyai hubungan linear baik terhadap radiasi matahari.

Pengujian Model

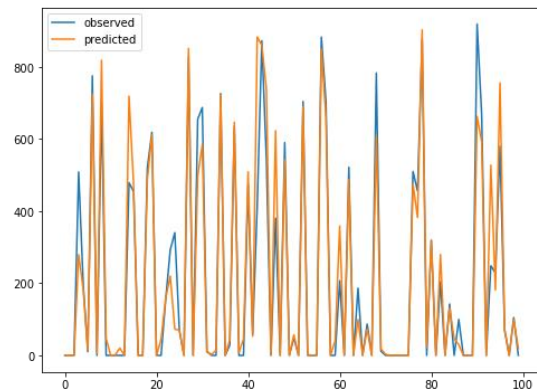
Pengujian model menggunakan python 3.7 dengan beberapa library seperti numpy, pandas, scipy, matplotlib, scikit-learn (sklearn). Model yang akan di evaluasi menggunakan model *Linear regression, Random Forest Regressor*

dan *Decision Tree Regressor*. Proses yang dilakukan yaitu dari data set yang tersedia diekstrak fitur-fitur penting dan mengurangi dimensi data dengan membuang data yang error atau data yang tidak mempunyai nilai. Data yang terkumpul dibagi secara acak 80% digunakan untuk pelatihan, 20% untuk pengujian dan validasi. Data pengujian tidak berpengaruh pada pelatihan. Hal ini dapat memberikan ukuran independen kinerja model selama dan setelah pelatihan. Kinerja model akan dievaluasi menggunakan matriks statistik berdasarkan beberapa indikator yaitu MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), R Square. Dari model yang di hasilkan, nilai prediksi dibandingkan dengan nilai observasi aktual. Nilai prediksi radiasi matahari yang di hasilkan dengan menggunakan metode Linear Regression dapat dilihat pada gambar di bawah.

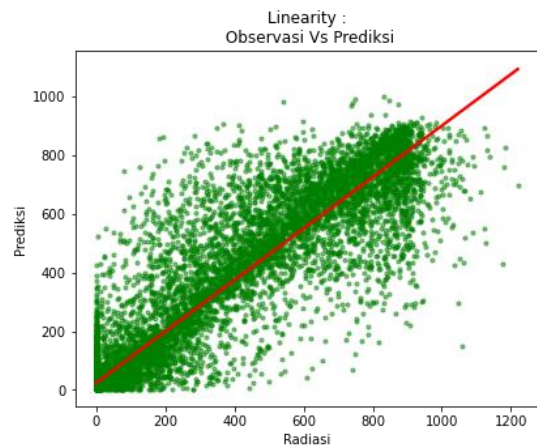
metode *random forest Regressor* dapat di lihat pada gambar di bawah.



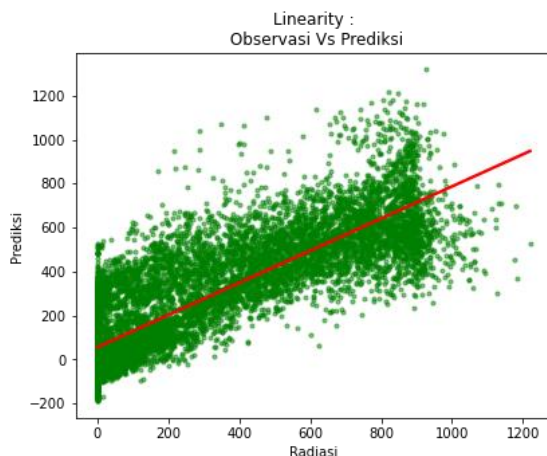
Gambar 8. Grafik Prediksi vs Observasi model linear regression



Gambar 10. Grafik Prediksi vs Observasi model random forest regressor



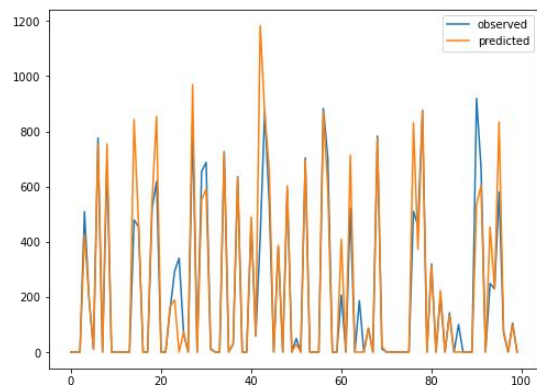
Gambar 11. Linearitas Prediksi vs Radiasi model random forest regressor



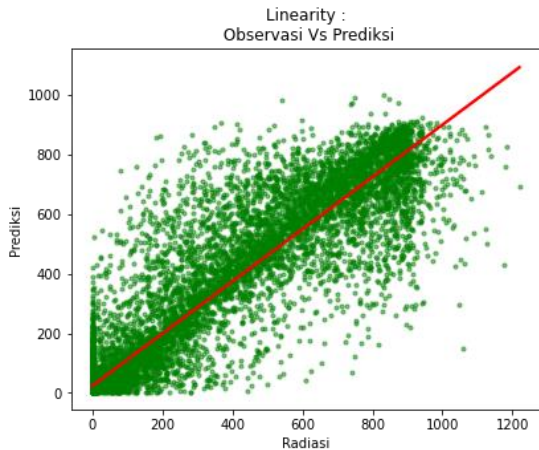
Gambar 9. Linearitas Prediksi vs Radiasi model linear regression

Hubungan antara nilai aktual dan nilai prediksi dengan grafik dari gambar 8 dan scatter plot dari gambar 9, mempunyai nilai MAE sebesar 110,5, nilai RMSE sebesar 154,8 dan R^2 mempunyai nilai 0,72. Prediksi radiasi matahari dengan

Dari gambar scatter plot dan grafik menggunakan metode random forest regressor di dapatkan hasil. Nilai MAE sebesar 51,4, nilai RMSE 106.0 dan nilai R^2 sebesar 0,87. Pengujian dengan metode *Decision Tree Regressor* dapat dilihat pada gambar di bawah.



Gambar 12. Grafik Prediksi vs Observasi model decision tree regressor



Gambar 13. Linearitas Prediksi vs Radiasi model decision tree regressor

Untuk pengujian menggunakan metode decision tree regressor yang dapat dilihat grafik dan scatter plotnya. Didapat nilai MAE sebesar 60,7 nilai RMSE sebesar 143,9 dan R^2 sebesar 0,76.

4. Kesimpulan

Dari ketiga metode prediksi yang di coba, menunjukkan bahwa prediksi dengan menggunakan metode *Random Forest Regressor* merupakan prediksi yang terbaik. Dimana nilai R^2 menunjukkan nilai 0.87. Nilai MAE dari ketiga metode menunjukkan nilai MAE terkecil yaitu 51,4 yang dihasilkan dari metode random forest regressor. Kemudian untuk nilai RMSE terkecil 106.0 dari metode *random forest regressor*. Secara keseluruhan metode random forest regressor merupakan metode yang lebih baik dari metode lainnya yang di ujikan di ikuti dengan metode decision tree dan kemudian linear regression.

Daftar Pustaka

- [1]. Buncom, F. (2017). *The Rise of Solar Energy*. Diakses tanggal 23 Agustus 2022 dari <http://large.stanford.edu/courses/2017/ph240/buncom1/>.
- [2]. Ramadhani, S, P. (2018). *Buku Bumi dan antariksa*. Depok, Jawa Barat : Yiesia Media Karya.
- [3]. Usvika, R. (2019). *Prospek Energi Surya dan Minihidro di Indonesia*. Di akses 24 Agustus 2022 dari <https://www.pupuk-indonesia.com/artikel/prospek-energi-surya-dan-mini-hidro-di-indonesia>.
- [4]. Eduardo et al (2020). Solar Radiation Prediction on Photovoltaic Systems Using Machine Learning Techniques. *Revista Facultad de Ingeniería.*, 29 (54). doi : 10.19053/01211129.v29.n54.2020.11751.
- [5]. Guher, A, B. Effective Estimation of Hourly Global Solar Radiation Using Machine Learning Algorithms. *International Journal of Photoenergy*. Vol. 2020, Article ID 8843620, <https://doi.org/10.1155/2020/8843620>
- [6]. Priya, S (2015). Solar Radiation Prediction Using Artificial Neural Network. *International Journal of Computer Application*. Vol. 116 – No. 16.
- [7]. Rahman et al (2021). Prediction of Solar Radiation Using Artificial Neural Network. *Journal of Physics: Conference Series*, 1767 012041. doi : 10.1088/1742-6596/1767/1/012041.
- [8]. Olatomiwa, L., Mekhilef, S., Shamshirband, S., & Petkovic, D. (2015). Adaptive neuro-fuzzy approach for solar radiation prediction in Nigeria. *Renewable and Sustainable Energy Reviews*, 51,1784-1791. doi : 10.1016/j.rser.2015.05.068.
- [9]. Voyant et al. (2017). Machine learning methods for solar radiation forecasting. *Renewable Energy*, 105, 569-582. doi : 10.1016/j.renene.2016.12.095.
- [10]. Mosavi, A., Ozturk, P., & Chau, K. (2018). Flood prediction using machine learning models. *MDPI*. 10(11), 1536. doi : 10.3390/w10111536.
- [11]. Huang et al (2021). Solar Radiation Prediction Using Different Machine Learning Algorithms and Implications for Extreme Climate Events. *Frontiers in Science 2021*, 9. doi : 10.3389/feart.2021.596860
- [12]. Priya, S. S., Iqbal, M. H. (2015). Solar Radiation Prediction using Artificial Neural Network. *International Journal of Computer Applications*, 116(16). doi : 10.5120/20422-2722.
- [13]. Rahman, S., Haque, B. (2020). Prediction of Solar Radiation Using Artificial Neural Network. *Journal of Physics: Conference Series*, 1767 (012041). doi : 10.1088/1742-6596/1767/1/012041.
- [14]. Wahyono, T. (2018). *Buku python for machine learning*. Yogyakarta : Gava Media.
- [15]. Trivusi (2023). *Algoritma Random Forest: Pengertian dan Kegunaannya*. Diakses dari <https://www.trivusi.web.id/2022/08/algoritma-random-forest.html>.
- [16]. Universitas Gajah Mada (2022). *Machine Learning*. Diakses dari <https://machinelearning.mipa.ugm.ac.id/2018/07/28/random-forest/>.
- [17]. Zhou, J., Gandomi, A., Chen, F.,& Holzinger, A. (2021). Evaluating the Quality of Machine Learning Explanations. *MDPI journal Electronics*, 105, 569-582. doi : 10.1016/j.renene.2016.12.095.