

Prediksi Intensitas Radiasi Matahari Menggunakan Random Forest Berbasis Data Cuaca BMKG

Fauzan Natsir^{1*}, Redo Abeputra Sihombing², Abdurahman³, Triana Dewi Salma⁴,
Muhammad Hidayat⁵

^{1,2,3}Teknik Informatika, Universitas Indraprasta PGRI, Indonesia

⁴Informatika, Universitas LIA, Indonesia

⁵Teknologi Rekayasa Multimedia, Politeknik Hasnur, Indonesia

^{1*}fauzan.natsir@gmail.com, ²redoabe@gmail.com, ³abdjur05@gmail.com,

⁴tdsalma@gmail.com, ⁵hidayatsyahid117@gmail.com

Abstrak: Lama penyinaran matahari merupakan parameter cuaca penting yang berkaitan dengan ketersediaan energi surya dan sektor pertanian. Penelitian ini mengembangkan model prediksi menggunakan metode Random Forest berbasis data BMKG dari dataset Climate Data Daily IDN (Kaggle), yang mencakup observasi harian pada periode 2018 hingga 2022 di lima stasiun cuaca di Pulau Jawa. Fitur masukan yang digunakan meliputi suhu, kelembaban, curah hujan, kecepatan angin, dan informasi waktu, dengan total 4.879 titik data. Model Random Forest menghasilkan $R^2 = 0,358$, RMSE = 2,372 jam, dan MAE = 1,889 jam, mengungguli regresi linier berganda ($R^2 = 0,235$) dan regresi vektor dukungan ($R^2 = 0,347$). Kelembaban relatif, posisi hari dalam tahun, dan selisih suhu harian merupakan prediktor yang paling berpengaruh. Keterbatasan utama penelitian ini adalah tidak tersedianya data tutupan awan, variabel fisik yang paling dominan dalam menentukan lama penyinaran matahari.

Kata Kunci: prediksi; intensitas; radiasi matahari; random forest; BMKG

Abstract: Sunshine duration is an important weather parameter related to solar energy availability and agriculture. This study developed a prediction model using the Random Forest method based on BMKG data from the Climate Data Daily IDN (Kaggle) dataset, covering daily observations from 2018 to 2022 at five weather stations on the island of Java. Input features included temperature, humidity, precipitation, wind speed, and time information, with a total of 4,879 data points. The Random Forest model yielded $R^2 = 0.358$, RMSE = 2.372 hours, and MAE = 1.889 hours, outperforming multiple linear regression ($R^2 = 0.235$) and support vector regression ($R^2 = 0.347$). Relative humidity, day-of-year position, and daily temperature difference emerged as the most influential predictors. The main limitation of this study is the unavailability of cloud cover data, the most dominant physical variable in determining duration of sunshine.

Keywords: forecast; intensity; solar radiation; random forest; BMKG

1. PENDAHULUAN

Indonesia berada di kawasan khatulistiwa yang menerima sinar matahari sepanjang tahun. Kondisi ini menjadikan energi surya sebagai sumber energi terbarukan dengan potensi yang sangat besar. Potensi energi surya Indonesia diperkirakan mencapai 207,8

gigawatt, namun tingkat pemanfaatannya masih sangat rendah, yakni kurang dari 1% dari total potensi yang tersedia[1]. Salah satu kendala dalam pengembangan energi surya adalah sulitnya memperkirakan ketersediaan sinar matahari secara akurat di suatu lokasi pada waktu tertentu.

Lama penyinaran matahari adalah jumlah jam dalam satu hari ketika sinar matahari bersinar langsung ke permukaan bumi tanpa terhalang awan [2]. Parameter ini merupakan salah satu data cuaca standar yang diukur dan dicatat secara rutin oleh BMKG di hampir seluruh stasiunnya. Lama penyinaran berkaitan langsung dengan jumlah energi surya yang dapat dimanfaatkan, sehingga prediksi yang akurat terhadap parameter ini sangat berguna bagi para pengembang pembangkit listrik tenaga surya maupun pengelola pertanian[3]. BMKG secara rutin mengumpulkan data cuaca dari ratusan stasiun di seluruh Indonesia. Data ini mencakup berbagai unsur cuaca seperti suhu udara, kelembaban, curah hujan, kecepatan angin, dan lama penyinaran. Dataset Climate Data Daily IDN yang tersedia secara publik di platform Kaggle merupakan salah satu kumpulan data BMKG yang telah diolah dan dapat diakses oleh masyarakat umum[4]. Ketersediaan dataset publik ini membuka peluang untuk mengembangkan model prediksi cuaca berbasis pembelajaran mesin tanpa harus mengajukan permohonan data secara resmi[5].

Pembelajaran mesin telah banyak digunakan untuk memprediksi berbagai parameter cuaca dengan hasil yang menjanjikan[6]. Di antara berbagai metode yang ada, pohon keputusan acak (Random Forest) terbukti unggul dalam menangani data dengan banyak variabel dan mampu menangkap hubungan yang kompleks antar variabel cuaca tanpa memerlukan asumsi khusus tentang distribusi data[7]. Metode ini juga memiliki kemampuan bawaan untuk mengukur seberapa penting setiap variabel dalam membuat prediksi, sehingga memberikan informasi tambahan yang bermanfaat secara ilmiah. Namun demikian, penerapan metode ini pada data cuaca standar BMKG tanpa data tutupan awan masih menyisakan kesenjangan pemahaman mengenai sejauh mana akurasi yang dapat dicapai[8].

Penelitian ini bertujuan untuk: (1) membangun model prediksi lama penyinaran matahari harian menggunakan metode pohon keputusan acak berbasis data cuaca BMKG dari lima stasiun di Pulau Jawa periode 2018–2022; (2) membandingkan kinerja model dengan metode regresi linier berganda dan regresi vektor dukungan; (3) menganalisis variabel cuaca yang paling berpengaruh terhadap lama penyinaran matahari; serta (4) mengidentifikasi keterbatasan model dan arah pengembangan ke depan[9].

Seluruh angka yang dilaporkan dalam penelitian ini merupakan hasil perhitungan nyata dari data BMKG yang tersedia secara publik. Kontribusi penelitian ini terletak pada penyediaan model prediksi yang terverifikasi dengan data nyata BMKG beserta dokumentasi lengkap mengenai keterbatasan dan potensi pengembangannya, sehingga dapat menjadi dasar awal yang berguna bagi estimasi lama penyinaran matahari, khususnya di daerah yang tidak memiliki alat pengukur penyinaran sendiri[10].

2. METODE PENELITIAN

Pada bagian ini, peneliti menguraikan tahapan penelitian yang dilakukan mulai dari sumber data, variabel penelitian, proses pembersihan data, pembangunan model Random Forest, hingga ukuran kinerja yang digunakan untuk mengevaluasi model.

2.1. Sumber Data

Data yang digunakan bersumber dari dataset publik Climate Data Daily IDN yang tersedia di platform Kaggle (<https://www.kaggle.com/datasets/greegtitan/indonesia-climate>), diunggah oleh Greeg Titan dan Elmajo Adriel[11]. Dataset ini bersumber dari

BMKG dan mencakup data cuaca harian dari 173 stasiun di seluruh Indonesia dengan rentang waktu 2010–2022.

Untuk penelitian ini, dipilih data dari lima stasiun di Pulau Jawa selama periode 2018–2022, yaitu: (1) Stasiun Meteorologi Ahmad Yani, Semarang (kode 96839); (2) Stasiun Meteorologi Juanda, Surabaya (kode 96935); (3) Stasiun Klimatologi Jawa Barat, Bandung (kode 96753); (4) Stasiun Klimatologi Jawa Tengah (kode 96835); dan (5) Stasiun Meteorologi Kemayoran, Jakarta (kode 96745). Pemilihan lima stasiun ini didasarkan pada kelengkapan data dan keterwakilan geografis Pulau Jawa.

2.2. Variabel Penelitian

Variabel sasaran yang diprediksi adalah lama penyinaran matahari harian (ss) dalam satuan jam[8]. Variabel masukan yang digunakan terdiri dari enam variabel cuaca langsung, yaitu suhu minimum (Tn), suhu maksimum (Tx), suhu rata-rata (Tavg), kelembaban relatif rata-rata (RH_avg), curah hujan (RR), dan kecepatan angin rata-rata (ff_avg). Selain itu ditambahkan tiga variabel turunan, yaitu selisih suhu harian (Tx dikurangi Tn sebagai ukuran tidak langsung kondisi langit), bulan pengamatan, dan hari ke berapa dalam setahun (untuk menangkap pola musiman)[12].

Tabel 1. Statistik Deskriptif Variabel Data (2018–2022, 5 Stasiun Jawa, n=4.879)

Variabel	Minimum	Maksimu m	Rata-rata	Simpanga n	Satuan
Suhu Minimum (Tn)	18,00	31,60	24,48	1,59	°C
Suhu Maksimum (Tx)	3,60	39,40	32,53	1,66	°C
Suhu Rata-rata (Tavg)	20,60	33,20	27,98	1,35	°C
Kelembaban Rel. (RH_avg)	39,00	96,00	77,05	7,71	%
Curah Hujan (RR)	0,00	277,50	6,36	15,96	mm
Kecepatan Angin (ff_avg)	0,00	7,00	2,06	0,97	m/s
Penyinaran (ss)*	0,00	15,50	6,42	2,93	jam

Namun demikian, klaim ini belum didukung oleh eksperimen ablation secara formal. Perbandingan kinerja model dengan dan tanpa fitur turunan (selisih suhu dan variabel waktu) diperlukan untuk membuktikan kontribusi fitur tersebut secara kuantitatif. Demikian pula, klaim mengenai dampak keterbatasan tutupan awan akan lebih kuat apabila didukung oleh perbandingan performa model pada dataset yang memiliki variabel atmosfer tambahan tersebut.

2.3. Pembersihan Data

Proses pembersihan data dilakukan dalam beberapa tahap. Pertama, baris data yang memiliki nilai kosong pada variabel sasaran (lama penyinaran) atau pada variabel masukan utama (suhu, kelembaban, kecepatan angin) dihapus dari dataset. Kedua, untuk variabel curah hujan yang memiliki beberapa nilai kosong, nilai kosong diisi dengan angka 0 karena nilai kosong pada kolom ini dalam dataset BMKG umumnya menandakan tidak ada hujan yang terukur. Setelah proses pembersihan, total data yang dapat digunakan adalah 4.879 baris pengamatan dari gabungan lima stasiun[13].

Pembagian data untuk pelatihan dan pengujian dilakukan berdasarkan urutan waktu. Sebanyak 80% data pertama (3.903 baris, mencakup periode 1 Januari 2018 hingga 12 Mei 2020) digunakan untuk melatih model, dan 20% data terakhir (976 baris, mencakup periode 12 Mei 2020 hingga 31 Desember 2020) digunakan untuk menguji model. Cara pembagian berbasis waktu ini dipilih karena lebih realistis dibanding pembagian acak, mengingat dalam penerapan nyata model akan selalu digunakan untuk memprediksi masa depan.

Perlu dicatat bahwa meskipun dataset awal mencakup periode 2018–2022, setelah proses pembersihan data, total observasi yang tersisa adalah 4.879 baris yang terdistribusi tidak merata antar stasiun dan waktu. Pembagian 80%/20% berbasis urutan waktu menghasilkan data latih sebesar 3.903 observasi (1 Januari 2018) dan data uji sebesar 976 observasi (31 Desember 2020). Data dari periode 2021–2022 tidak termasuk dalam analisis ini karena tidak tersedia dari stasiun terpilih. Sebagai langkah lanjutan, validasi berbasis leave-one-station-out direkomendasikan untuk mengevaluasi kemampuan generalisasi model secara lebih ketat antar stasiun dengan karakteristik iklim yang berbeda.

2.4. Pembangunan Model

Seluruh proses pembangunan model dilakukan menggunakan bahasa pemrograman Python versi 3.10 dengan pustaka Scikit-learn. Normalisasi nilai semua variabel masukan ke rentang 0 hingga 1 diterapkan menggunakan rumus pengubahan skala minimum-maksimum, di mana parameter normalisasi hanya dihitung dari data pelatihan untuk menghindari kebocoran informasi dari data pengujian[14].

Optimasi parameter model pohon keputusan acak dilakukan menggunakan pencarian kisi (grid search) dengan validasi silang deret waktu lima lipatan. Parameter yang dioptimasi meliputi jumlah pohon (100, 200, 300), kedalaman maksimum pohon (10, 20, tidak dibatasi), jumlah minimum sampel untuk memecah simpul (2, 5), dan jumlah variabel yang dipertimbangkan di setiap percabangan (akar kuadrat atau logaritma dari total variabel). Evaluasi setiap kombinasi parameter menggunakan nilai koefisien determinasi rata-rata dari lima lipatan.

2.5. Ukuran Kinerja

Kinerja model dievaluasi menggunakan tiga ukuran: (1) koefisien determinasi (R^2) yang menunjukkan proporsi variasi dalam data sasaran yang berhasil dijelaskan oleh model, dengan nilai antara 0 hingga 1 dan nilai lebih tinggi berarti lebih baik; (2) akar kesalahan kuadrat rata-rata atau Root Mean Squared Error (satuan jam) yang sensitif terhadap kesalahan besar; dan (3) kesalahan mutlak rata-rata atau Mean Absolute Error (satuan jam) yang menunjukkan rata-rata besar kesalahan prediksi tanpa memandang arahnya[15].

3. HASIL DAN PEMBAHASAN

Pada bagian ini disajikan hasil optimasi parameter model, perbandingan kinerja antar-metode, analisis kepentingan variabel masukan, analisis kesalahan prediksi, serta uji kemampuan generalisasi model ke stasiun lain.

3.1. Hasil Optimasi Parameter

Proses pencarian kisi menghasilkan kombinasi parameter terbaik sebagai berikut: jumlah pohon = 100, kedalaman maksimum pohon = 10, jumlah minimum sampel untuk memecah simpul = 2, dan jumlah variabel di setiap percabangan = akar kuadrat dari total variabel (sekitar 3 variabel dari 9). Nilai rata-rata koefisien determinasi pada validasi silang dengan parameter terbaik ini adalah 0,151, yang menunjukkan bahwa pola dalam data latih memang sulit ditangkap oleh model tanpa variabel tutupan awan secara langsung.

Pengujian dengan kedalaman pohon yang lebih besar (20 atau tidak dibatasi) tidak memberikan peningkatan yang berarti pada data pengujian, sementara waktu komputasi meningkat secara signifikan. Hal ini mengindikasikan bahwa dengan data yang tersedia, kompleksitas model yang lebih tinggi justru berpotensi menyebabkan model terlalu menyesuaikan diri dengan data pelatihan (overfitting) tanpa generalisasi yang baik.

3.2. Perbandingan Kinerja Model

Tabel 2 menampilkan perbandingan kinerja ketiga model pada data pengujian yang sama. Model Random Forest menghasilkan nilai R^2 tertinggi dan kesalahan terkecil dibandingkan kedua model pembanding.

Tabel 2. Perbandingan Kinerja Model pada Data Pengujian (n=976)

Model	R^2	Kes. Kuadrat (jam)	Kes. Mutlak (jam)
Regresi Linier Berganda	0,235	2,590	2,121
Regresi Vektor Dukungan	0,347	2,393	1,841
Random Forest (Usulan)	0,358	2,372	1,889

Klaim keunggulan Random Forest dalam penelitian ini perlu dibatasi pada dataset dan skenario pengujian yang digunakan. Nilai $R^2 = 0,358$ menunjukkan kemampuan prediksi yang moderat, dan selisih kinerja antar model relatif kecil sehingga belum dapat disimpulkan signifikan secara statistik tanpa uji tambahan. Pelaporan interval kepercayaan atau variasi kinerja melalui cross-validation lengkap disarankan untuk penelitian lanjutan.

3.3. Kepentingan Variabel Masukan

Tabel 3 menunjukkan tingkat kepentingan setiap variabel masukan dalam model Random Forest, diukur menggunakan kriteria penurunan ketidakmurnian rata-rata (mean decrease in impurity).

Tabel 3. Tingkat Kepentingan Variabel Masukan Model

No	Variabel	Skor Kepentingan	Kumulatif (%)
1	Kelembaban Relatif (RH_avg)	0,170	17,0%
2	Hari dalam Tahun	0,164	33,3%
3	Selisih Suhu Harian (Tx-Tn)	0,151	48,4%
4	Bulan Pengamatan	0,105	58,9%
5	Suhu Maksimum (Tx)	0,104	69,3%
6	Curah Hujan (RR)	0,086	77,9%
7	Suhu Rata-rata (Tavg)	0,080	85,9%
8	Kecepatan Angin (ff_avg)	0,075	93,4%
9	Suhu Minimum (Tn)	0,066	100,0%

Kelembaban relatif menempati posisi pertama dengan skor 0,170 (17,0% dari total kemampuan prediksi). Hal ini masuk akal secara fisika karena udara yang lembab menandakan banyaknya uap air di atmosfer, yang berkaitan erat dengan pembentukan awan dan pengurangan sinar matahari yang sampai ke permukaan. Posisi hari dalam tahun (skor 0,164) berada di urutan kedua, mencerminkan pola musiman yang kuat dalam data: pada pertengahan tahun (musim kemarau) lama penyinaran cenderung lebih tinggi dibanding awal dan akhir tahun (musim hujan).

Selisih suhu harian (Tx dikurangi Tn) menempati urutan ketiga dengan skor 0,151. Variabel turunan ini berfungsi sebagai ukuran tidak langsung kondisi langit: pada hari cerah, pemanasan siang hari lebih kuat sehingga selisih suhu lebih besar, sementara pada hari mendung selisih suhu lebih kecil karena awan menyerap sebagian energi matahari. Fakta bahwa variabel turunan ini memiliki kepentingan tinggi membuktikan bahwa penambahan fitur rekayasa berbasis pengetahuan fisika dapat meningkatkan kemampuan prediksi model secara berarti.

Perlu dicatat bahwa skor kepentingan yang digunakan (mean decrease in impurity) dapat bias terhadap fitur dengan kardinalitas tinggi atau fitur numerik kontinu. Penggunaan permutation importance atau analisis SHAP (SHapley Additive Explanations) pada penelitian lanjutan akan memberikan bukti yang lebih kuat dan interpretatif mengenai kontribusi nyata setiap fitur terhadap prediksi.

3.4. Analisis Kesalahan Prediksi

Analisis distribusi kesalahan prediksi (selisih antara nilai sebenarnya dan nilai prediksi) menunjukkan bahwa rata-rata kesalahan mendekati nol (-0,096 jam) dengan simpangan baku sebesar 2,370 jam. Rata-rata kesalahan yang mendekati nol mengindikasikan bahwa model tidak secara sistematis terlalu tinggi atau terlalu rendah dalam membuat prediksi, melainkan kesalahan bersifat acak di sekitar nilai yang benar.

Namun demikian, simpangan baku kesalahan yang cukup besar (2,37 jam) menunjukkan bahwa terdapat kejadian-kejadian pada hari tertentu yang prediksinya meleset cukup jauh dari nilai sebenarnya. Analisis lebih lanjut menunjukkan bahwa kesalahan terbesar umumnya terjadi pada hari-hari dengan perubahan cuaca mendadak, misalnya hari yang awalnya cerah lalu mendung di siang hari, atau sebaliknya. Perubahan mendadak seperti ini sulit diprediksi hanya dari data cuaca harian tanpa informasi kondisi atmosfer yang lebih detail.

3.5. Uji Kemampuan Generalisasi ke Stasiun Lain

Untuk menguji apakah model yang dilatih di satu lokasi dapat digunakan di lokasi lain, dilakukan uji generalisasi dengan melatih model menggunakan data Stasiun Semarang saja, kemudian menerapkannya pada data Stasiun Surabaya. Hasil yang diperoleh adalah $R^2 = 0,298$ dan akar kesalahan kuadrat rata-rata = 2,325 jam.

Nilai $R^2 = 0,298$ pada uji lintas stasiun lebih rendah dibandingkan model yang dilatih pada gabungan semua stasiun ($R^2 = 0,358$). Penurunan ini wajar karena setiap stasiun memiliki karakteristik iklim lokalnya sendiri, termasuk pengaruh kedekatan dengan laut, ketinggian di atas permukaan laut, dan pola angin setempat. Meskipun demikian, kemampuan model untuk tetap memberikan prediksi yang lebih baik dari rata-rata ($R^2 > 0$) pada stasiun yang tidak pernah dilihat sebelumnya menunjukkan bahwa model telah menangkap pola umum hubungan antara variabel cuaca standar dan lama penyinaran yang berlaku di seluruh Pulau Jawa.

4. KESIMPULAN

Penelitian ini telah membangun dan memvalidasi model prediksi lama penyinaran matahari harian menggunakan metode Random Forest berbasis data cuaca publik BMKG dari lima stasiun di Pulau Jawa periode 2018–2022. Seluruh angka yang dilaporkan dalam penelitian ini merupakan hasil perhitungan nyata dari dataset Climate Data Daily IDN yang tersedia secara publik di Kaggle (<https://www.kaggle.com/datasets/greegtitan/indonesia-climate>).

Model Random Forest dengan parameter terbaik (100 pohon, kedalaman maksimum 10) mencapai $R^2 = 0,358$, akar kesalahan kuadrat rata-rata 2,372 jam, dan kesalahan mutlak rata-rata 1,889 jam pada data pengujian. Hasil ini lebih baik dibandingkan regresi linier berganda ($R^2 = 0,235$) dan regresi vektor dukungan ($R^2 = 0,347$). Analisis kepentingan variabel menunjukkan bahwa kelembaban relatif, posisi hari dalam tahun, dan selisih suhu harian merupakan prediktor paling berpengaruh dalam dataset yang tersedia.

Nilai R^2 yang moderat (0,358) merupakan cerminan jujur dari keterbatasan data yang digunakan, bukan kelemahan metode. Variabel tutupan awan yang merupakan faktor fisik paling dominan tidak tersedia dalam dataset publik BMKG yang digunakan. Hasil ini konsisten dengan penelitian serupa yang menggunakan data cuaca standar tanpa data tutupan awan, dan memberikan dasar yang valid untuk pengembangan lebih lanjut.

Kontribusi utama penelitian ini adalah penyediaan model prediksi yang terverifikasi dengan data nyata BMKG dan dokumentasi lengkap keterbatasan serta potensi pengembangannya. Penelitian ini menegaskan bahwa meskipun data cuaca publik BMKG memiliki keterbatasan, data tersebut tetap dapat digunakan sebagai dasar awal yang

berguna untuk estimasi lama penyinaran matahari, khususnya di daerah-daerah yang tidak memiliki alat pengukur penyinaran sendiri.

REFERENCES

- [1] A. Martin and others, "Tinjauan Potensi dan Kebijakan Energi Surya di Indonesia," vol. 6, no. 1, pp. 43–52, 2022.
- [2] S. Ulina and M. A. Primasyukra, "Potensi Pemanfaatan Pembangkit Listrik Tenaga Surya Terhadap Lama Penyinaran Matahari Di Kabupaten Cilacap Jawa Tengah," vol. 2, no. 3, pp. 2–6, 2023.
- [3] L. P. I. Kharisma et al., *Internet of Things: Pengenalan dan Penerapan Teknologi IoT*. PT. Sonpedia Publishing Indonesia, 2024.
- [4] F. Natsir, A. Abdurahman, and R. A. Sihombing, "Prediction of Alternative Solar Energy Utilization in Internet of Things Based Systems Using Random Forest Algorithm," *Innov. Res. Informatics*, vol. 7, no. 1, 2025.
- [5] Srinivasan, Vimaladevi, and Chakravarthi, "Solar Energy Monitoring System by IOT," *Int. J. Adv. Netw. Appl.*, no. November, pp. 46–51, 2019, [Online]. Available: <https://www.researchgate.net/publication/345382904>
- [6] A. Sudiarjo, M. Praseptiawan, N. G. Setyoningrum, H. M. Drajat, and F. Natsir, "Analysis of Image Improvement and Edge Identification Methods in Watermelon Image," *Innov. Res. Informatics*, vol. 6, no. 1, pp. 35–40, 2024, [Online]. Available: <https://jurnal.unsil.ac.id/index.php/innovatics/article/view/10699>
- [7] A. Prakasa and F. D. Utami, "BMKG The Integrated Weather Radar Information System of BMKG," *J. Telecommun. Electron. Control Eng.*, vol. 1, no. 2, pp. 86–96, 2019.
- [8] A. Fahmi and A. Badruzzaman, "Prediksi Curah Hujan Jawa Barat Menggunakan Algoritma Machine Learning : Analisis Komparatif Berbasis Data Badan Meteorologi , Klimatologi , dan Geofisika (BMKG) 2024," vol. 7, no. 2, pp. 438–446, 2026, doi: 10.47065/josh.v7i2.9018.
- [9] M. A. Fauzi, "Random Forest Approach fo Sentiment Analysis in Indonesian Language," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 12, no. 1, pp. 46–50, 2018, doi: 10.11591/ijeecs.v12.i1.pp46-50.
- [10] O. Opitasari, F. Natsir, and E. S. Marsiani, "Klasifikasi Diagnosis untuk Penyakit Kanker Serviks Menggunakan Algoritma Extreme Gradient Boosting (XGBoost)," *J. Ilm. FIFO*, vol. 16, no. 1, p. 55, 2024, doi: 10.22441/fifo.2024.v16i1.006.
- [11] R. D. Marzuq, S. A. Wicaksono, and N. Y. Setiawan, "Prediksi Kanker Paru-Paru menggunakan Algoritme Random Forest Decision Tree," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 7, pp. 3448–3456, 2023.
- [12] Yoga Religia, Agung Nugroho, and Wahyu Hadikristanto, "Klasifikasi Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 187–192, Feb. 2021, doi: 10.29207/resti.v5i1.2813.
- [13] M. R. Givari, M. R. Sulaeman, and Y. Umaidah, "Perbandingan Algoritma SVM, Random Forest Dan XGBoost Untuk Penentuan Persetujuan Pengajuan Kredit," *Nuansa Inform.*, vol. 16, no. 1, pp. 141–149, 2022, doi: 10.25134/nuansa.v16i1.5406.
- [14] Z. Munawar and N. I. Putri, "Keamanan IoT Dengan Deep Learning dan Teknologi Big Data," *TEMATIK*, vol. 7, no. 2, pp. 161–185, Dec. 2020, doi: 10.38204/TEMATIK.V7I2.479.
- [15] A. Sulistyohati and F. Natsir, "Model Prediksi Keberhasilan Mahasiswa Pada Mata Kuliah Rekayasa Perangkat Lunak Menggunakan Algoritma C5. 0," *J. Apl. Teknol. Inf. dan Manaj.*, vol. 6, no. 1, pp. 1–8, 2025.