



Penerapan Algoritma *Neural Network* Untuk Prediksi Hujan Harian Di Kota Tegal Berbasis Data Klimatologi Nasa Power

Application of Neural Network Algorithms for Daily Rainfall Prediction in Tegal City Based on NASA Power Climatology Data

Ziamul Umam

Universitas Pancasakti Tegal

Email: ziamulumamm@gmail.com

Article Info

Article history :

Received : 28-12-2025

Revised : 30-12-2025

Accepted : 02-01-2026

Pulished : 04-01-2026

Abstract

Tegal City, located on the northern coast, is vulnerable to hydrometeorological risks, making accurate weather forecasting crucial. However, the limitation and discontinuity of ground observation data often hinder reliable prediction modeling. This study aims to predict daily rainfall using a Machine Learning approach based on long-term satellite data. Daily climatological data sourced from NASA POWER spanning from 2000 to 2025, totaling 9,466 records, were used as input. The method employed is the Artificial Neural Network (ANN) with the Backpropagation algorithm, processed using RapidMiner software. Given the time-series nature of the data, model validation was conducted using sequential Split Validation techniques. Test results indicate the model has good classification performance with an AUC value of 0.828. Although the overall accuracy is 77.71%, the model demonstrates very high sensitivity in detecting rainfall events, achieving a Recall of 88.82%. In conclusion, the ANN model is effective as an early detection instrument, where the ability to minimize missed rainfall detection (*False Negative*) is prioritized over precision in the context of disaster mitigation.

Keywords: *artificial neural network, rainfall prediction, weather forecasting.*

Abstrak

Kota Tegal sebagai wilayah pesisir utara rentan terhadap risiko bencana hidrometeorologi, sehingga informasi prakiraan cuaca yang akurat sangat krusial. Namun, keterbatasan dan ketidakkontinuan data observasi permukaan seringkali menghambat pemodelan prediksi yang handal. Penelitian ini bertujuan untuk memprediksi curah hujan harian menggunakan pendekatan *Machine Learning* berbasis data satelit jangka panjang. Data klimatologi harian bersumber dari NASA POWER periode tahun 2000 hingga 2025 dengan total 9.466 data digunakan sebagai input. Metode yang digunakan adalah *Artificial Neural Network* (ANN) dengan algoritma *Backpropagation* yang diproses menggunakan perangkat lunak RapidMiner. Mengingat karakteristik data berupa deret waktu (*time-series*), validasi model dilakukan menggunakan teknik *Split Validation* secara sekuensial. Hasil pengujian menunjukkan model memiliki performa klasifikasi yang baik dengan nilai AUC 0.828. Meskipun akurasi keseluruhan tercatat sebesar 77,71%, model menunjukkan sensitivitas yang sangat tinggi dalam mendeteksi kejadian hujan dengan nilai *Recall* mencapai 88,82%. Kesimpulannya, model ANN efektif digunakan sebagai instrumen deteksi dini, di mana kemampuan meminimalkan kegagalan deteksi hujan (*False Negative*) lebih diutamakan dibandingkan presisi semata dalam konteks mitigasi bencana.

Kata kunci: *artificial neural network, curah hujan, prediksi cuaca*



PENDAHULUAN

Kota Tegal merupakan wilayah pesisir yang memiliki kerentanan tinggi terhadap bencana banjir akibat perubahan pola cuaca yang tidak menentu. Fenomena ini menuntut adanya sistem peringatan dini yang efektif, sebagaimana ditekankan oleh Usman (2023) yang menyatakan bahwa sistem peringatan dini berbasis teknologi sangat krusial untuk mengurangi risiko korban dan kerugian materi akibat bencana hidrometeorologi [1]. Namun, upaya prediksi seringkali terkendala oleh karakteristik hujan yang sangat dinamis dan dipengaruhi oleh banyak faktor kompleks, membuat metode konvensional seringkali kurang akurat dalam menangkap pola perubahannya [2].

Masalah lain yang sering dihadapi adalah keterbatasan data observasi dari stasiun penakar hujan konvensional yang kerap mengalami kerusakan alat atau kekosongan data (missing values). Ketidaklengkapan data ini menghambat akurasi peramalan cuaca. Sebagai solusi, penggunaan data satelit global seperti NASA Prediction Of Worldwide Energy Resources (NASA POWER) telah terbukti mampu menjadi sumber data alternatif yang andal untuk melengkapi kekosongan data stasiun darat, terutama untuk estimasi hidrologi di wilayah yang minim instrumen [3].

Untuk mengolah data atmosfer yang memiliki karakteristik fluktuatif dan non-linear, metode Machine Learning khususnya Artificial Neural Network (ANN) terbukti memiliki performa yang lebih unggul. Studi terbaru oleh Nurmala, Fitri, dan Gautami (2024) menunjukkan bahwa teknik Backpropagation pada Neural Network mampu memetakan pola curah hujan yang kompleks dengan tingkat akurasi yang dapat diandalkan untuk keperluan mitigasi [4]. Penelitian ini bertujuan menerapkan algoritma ANN untuk memprediksi kejadian hujan harian di Kota Tegal menggunakan data runtun waktu jangka panjang (2000–2025). Kebaruan (novelty) penelitian ini terletak pada fokus evaluasi terhadap metrik Recall (Sensitivitas) guna meminimalkan kegagalan deteksi hujan (False Negative).

METODE PENELITIAN

Penelitian ini menerapkan kerangka kerja Knowledge Discovery in Databases (KDD) untuk membangun model prediksi cuaca. Tahapan penelitian meliputi pengumpulan data klimatologi, pra-pemrosesan data (preprocessing), pelatihan model Artificial Neural Network (ANN), dan evaluasi kinerja model.

Sumber dan Variabel Data

Data yang digunakan bersumber dari NASA POWER dengan koordinat lokasi Kota Tegal. Dataset mencakup periode pengamatan harian selama 26 tahun (2000-2025). Penggunaan dataset reanalisis seperti NASA POWER dan ECMWF telah diverifikasi memiliki korelasi yang baik terhadap kondisi riil di wilayah kepulauan Indonesia, sehingga layak digunakan untuk analisis potensi iklim dan cuaca [5]. Variabel input (X) terdiri dari parameter fisika atmosfer harian seperti temperatur, kelembapan, dan kecepatan angin, sedangkan variabel target (Y) adalah status kejadian hujan (H+1).



Tabel 1. Atribut Dataset Penelitian

Atribut	Deskripsi	Satuan
T2M	Temperatur Rata-rata Harian	°C
T2M_MAX	Temperatur Maksimum Harian	°C
T2M_MIN	Temperatur Minimum Harian	°C
RH2M	Kelembapan Relatif pada 2 Meter	%
WS2M	Kecepatan Angin pada 2 Meter	m/s
IMERG_PRECTOT	Curah Hujan Harian (Data Satelit)	mm
RAIN_TOMORROW	Status Hujan Besok (Target)	Binomial

Pra-pemrosesan Data

Tahap ini krusial untuk memastikan kualitas data. Data curah hujan numerik dikonversi menjadi label biner (Hujan/Tidak Hujan). Selanjutnya, dilakukan normalisasi Min-Max karena algoritma ANN sangat sensitif terhadap perbedaan skala data. Teknik preprocessing yang tepat, termasuk penanganan data time-series dengan metode sliding window atau sekuensial, sangat menentukan keberhasilan model dalam mengenali pola musiman [6].

1. Transformasi Label (Labeling)

Data curah hujan (IMERG_PRECTOT) yang bersifat numerik kontinu dikonversi menjadi data nominal biner (Hujan/Tidak Hujan) menggunakan fungsi thresholding. Mengingat sensitivitas sensor satelit, ambang batas hujan ditetapkan pada >1 mm untuk memfilter noise atau gerimis ringan yang seringkali tidak terukur di permukaan. Secara matematis didefinisikan sebagai:

$$Y_{(t+1)} = \begin{cases} 1(\text{Hujan}), & \text{jika } IMERG_PRECTOT > 1\text{mm} \\ 0(\text{TidakHujan}), & \text{jika } IMERG_PRECTOT \leq 1\text{mm} \end{cases}$$

Dimana Y_{t+1} adalah label target yang akan diprediksi berdasarkan fitur input hari ini (X_t).

2. Normalisasi Data

Mengingat algoritma ANN sensitif terhadap perbedaan skala antar atribut (misalnya Tekanan ~ 100 kPa vs Curah Hujan $\sim 0-50$ mm), dilakukan normalisasi menggunakan metode Min-Max Normalization untuk mentransformasi seluruh fitur input ke dalam rentang $[0, 1]$. Persamaan normalisasi adalah:

$$x' = \left(\frac{x - \min(x)}{\max(x) - \min(x)} \right)$$

3. Pembagian Data (Data Splitting)

Untuk menjaga kaidah data time-series, pembagian data latih (training) dan data uji (testing) dilakukan secara sekuensial atau berurutan berdasarkan waktu, bukan secara acak (random shuffle). Strategi ini bertujuan untuk mencegah kebocoran data (data leakage) dari informasi masa depan ke masa lalu, dengan alokasi data latih sebanyak 7.573 data yang diambil dari periode awal dan data uji sebanyak 1.893 data yang berasal dari periode akhir atau terbaru.



Arsitektur Model

Penelitian ini menggunakan algoritma Feed-forward Neural Network dengan metode pembelajaran Backpropagation. Metode ini dipilih karena kemampuannya meminimalkan error melalui pembaruan bobot secara iteratif. Sebagaimana dijelaskan dalam penelitian Utami dan Hidayatullah (2024), algoritma Backpropagation dengan konfigurasi layer dan learning rate yang tepat mampu menghasilkan akurasi tinggi (mencapai 95% pada kasus tertentu) dalam prediksi parameter cuaca [7].

Parameter hyperparameter yang digunakan dalam proses pelatihan dikonfigurasi sebagai berikut untuk mencapai konvergensi optimal:

1. Training Cycles (Epochs): 200
2. Learning Rate (η): 0.01
3. Momentum (α): 0.9
4. Error Epsilon: 1.0×10^{-4}
5. Hidden Layers: 1 Layer

Proses pembelajaran ANN bertujuan meminimalkan fungsi kerugian (loss function) dengan memperbarui bobot sinaptik (w) pada setiap iterasi berdasarkan error yang dihasilkan.

Evaluasi Model

Kinerja model dievaluasi menggunakan Confusion Matrix. Metrik utama yang menjadi fokus penelitian ini adalah Recall (Sensitivitas) dan Area Under Curve (AUC), di samping Accuracy dan Precision. Mengingat tujuan penelitian untuk peringatan dini, Recall menjadi prioritas untuk meminimalkan False Negative (kegagalan memprediksi hujan). Pemilihan rasio pembagian data yang tepat sangat berpengaruh terhadap validitas model Machine Learning; rasio yang tidak seimbang dapat menyebabkan bias pada hasil evaluasi performa [8]. Metrik dihitung dengan persamaan:

$$Accuracy = \frac{TP + TN}{FP + FN + TP + TN}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Dimana TP adalah True Positive, TN adalah True Negative, FP adalah False Positive, dan FN adalah False Negative. Nilai AUC digunakan untuk mengukur kemampuan model dalam membedakan kelas secara keseluruhan pada berbagai ambang batas klasifikasi.

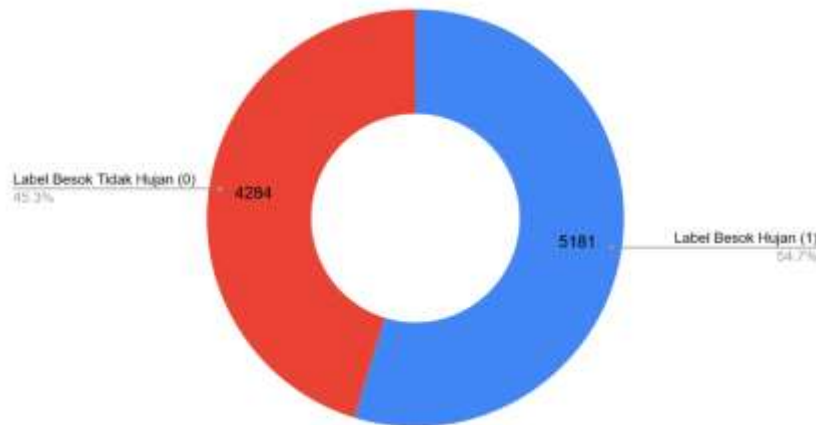
HASIL DAN PEMBAHASAN

Bab ini menyajikan analisis komprehensif mengenai kinerja model Artificial Neural Network (ANN) yang telah dilatih untuk memprediksi curah hujan harian di Kota Tegal. Eksperimen dilakukan menggunakan dataset time-series dari NASA POWER periode 2000–2025, dengan fokus pengujian pada data masa depan (tahun 2020–2025) yang belum pernah dilihat oleh model selama fase pelatihan (unseen data).



Distribusi dan Kualitas Data

Sebelum melangkah pada evaluasi model, penting untuk meninjau karakteristik statistik dari dataset yang digunakan. Kualitas distribusi kelas target sangat menentukan perilaku model ANN dalam mengambil keputusan.



Gambar 1. Proporsi Distribusi Kelas Target (Hujan vs Tidak Hujan)

Dataset menunjukkan karakteristik yang sangat ideal untuk pembelajaran mesin. Dari total 9.466 sampel data, terdapat 5.181 kejadian hujan (54,7%) dan 4.284 kejadian tidak hujan (45,3%). Distribusi ini tergolong seimbang (balanced). Temuan ini menarik karena pada umumnya data cuaca di wilayah tropis memiliki kecenderungan ketidakseimbangan kelas (imbalanced), di mana hari kering seringkali mendominasi. Keseimbangan alami pada data NASA POWER ini memberikan keuntungan signifikan secara metodologis, yaitu menghilangkan bias prediksi. Model ANN tidak akan terjebak dalam kondisi overfitting terhadap kelas mayoritas, sehingga tidak diperlukan teknik augmentasi data sintetik seperti SMOTE (Synthetic Minority Over-sampling Technique) yang berpotensi memunculkan noise buatan. Keseimbangan ini menjadi fondasi awal yang kuat bagi model untuk mengenali pola hujan secara objektif.

Evaluasi Kuantitatif Menggunakan Confusion Matrix

Untuk mengukur akurasi prediksi secara detail, model diuji menggunakan 1.893 data uji. Hasil klasifikasi dipetakan ke dalam Confusion Matrix sebagaimana ditampilkan pada Tabel 2. Matriks ini memberikan gambaran granular mengenai di mana letak kesalahan dan kebenaran prediksi model.

Tabel 2. Confusion Matrix

	true 1	true 0	class precision
pred. 1	985	298	76.77%
pred. 0	124	486	79.67%
class recall	88.82%	61.99%	



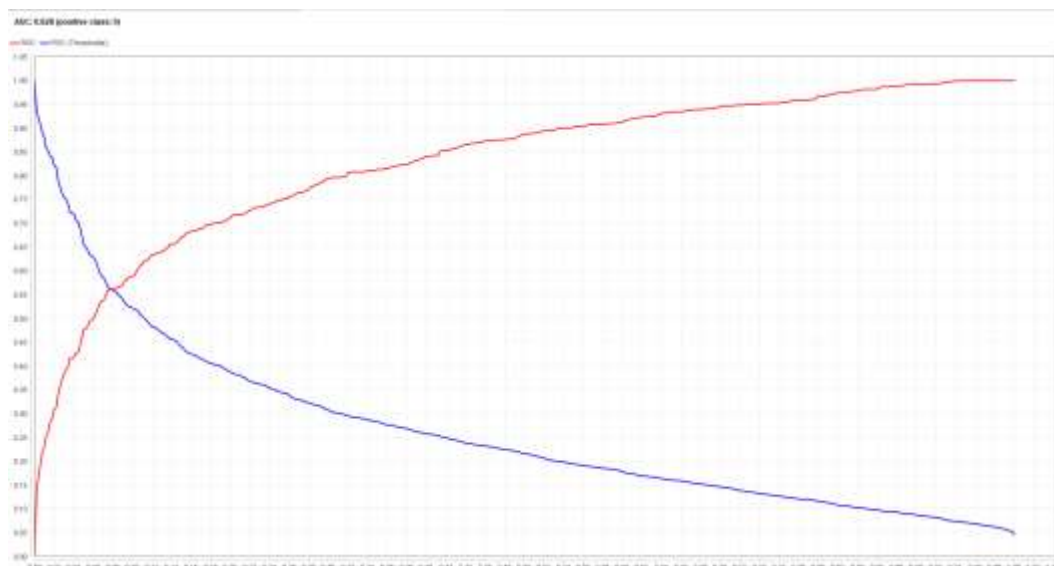
Berdasarkan Tabel 2, dapat dianalisis beberapa poin krusial:

1. Dominasi True Positive: Model berhasil memprediksi dengan tepat 985 kejadian hujan. Angka ini menunjukkan bahwa ANN mampu menangkap pola sinoptik (suhu, kelembapan, tekanan) yang menjadi prekursor turunnya hujan dengan sangat baik.
2. False Positive (Over-estimation): Terdapat 298 kasus di mana model memprediksi hujan, namun kenyataannya tidak hujan. Kesalahan ini mengindikasikan bahwa model cenderung "agresif" atau sensitif. Secara meteorologis, hal ini dapat terjadi ketika kondisi atmosfer sudah jenuh (mendung tebal), namun hujan tidak terukur di permukaan tanah (misalnya karena fenomena virga atau evaporasi).
3. False Negative (Missed Detection): Kesalahan fatal di mana hujan terjadi namun model gagal memprediksi hanya berjumlah 124 kasus. Rendahnya angka ini sangat positif untuk tujuan mitigasi bencana.

Secara keseluruhan, model mencapai Akurasi 77.71%. Angka ini menunjukkan konsistensi model yang baik, mengingat cuaca adalah sistem stokastik yang sangat dinamis dan sulit diprediksi secara linear. Hasil ini dapat dibandingkan dengan penelitian Utami & Hidayatullah (2024) yang juga memprioritaskan Recall (97.5%) untuk memastikan tidak ada kejadian hujan ekstrem yang terlewatkan dalam prediksi [7]. Meskipun akurasi total penelitian ini (77%) sedikit di bawah beberapa studi lain yang mencapai 80-90%, namun nilai Recall yang tinggi menunjukkan model ini lebih "aman" untuk digunakan dalam mitigasi bencana.

Analisis Kemampuan Diskriminasi Model (AUC-ROC)

Selain akurasi titik, stabilitas model pada berbagai ambang batas (threshold) diuji menggunakan kurva Receiver Operating Characteristic (ROC).



Gambar 2. Kurva ROC dan Nilai AUC Model

Kurva pada Gambar 3 memperlihatkan garis yang melengkung tajam mendekati sudut kiri atas, yang merepresentasikan performa klasifikasi yang ideal. Luas area di bawah kurva (Area Under Curve - AUC) tercatat sebesar 0.828.



Nilai AUC 0.828 masuk dalam kategori "Excellent Classification". Hal ini bermakna bahwa model ANN memiliki probabilitas 82,8% untuk membedakan antara hari hujan dan hari kering dengan benar. Nilai ini mengonfirmasi bahwa fitur-fitur input yang dipilih (T2M, RH2M, WS2M, PS) memiliki korelasi yang kuat terhadap target prediksi, dan arsitektur ANN yang dibangun berhasil memisahkan kedua kelas tersebut secara efektif di ruang dimensi fitur. Temuan ini mendukung kesimpulan Sanidhya et al. (2023) bahwa meskipun data satelit memiliki bias tertentu, informasinya cukup kaya untuk memodelkan pola hidrologi di wilayah yang minim data observasi [3].

Prioritas Recall untuk Sistem Peringatan Dini

Bagian terpenting dari penelitian ini bukan sekadar pada tingginya Akurasi, melainkan pada profil performa spesifik model dalam konteks aplikatif. Tabel 3 menyajikan perbandingan metrik performa secara komprehensif.

Tabel 3. Perbandingan Metrik Performa Model

Metrik Evaluasi	Nilai Capaian (%)
Accuracy	77.71%
AUC	0.828
Precision (Class 1: Hujan)	76.77%
Recall (Class 1: Hujan)	88.82%
F1-Score	82.35%

Analisis mendalam terhadap metrik di atas menghasilkan temuan signifikan:

1. Keunggulan Sensitivitas (Recall)

Nilai Recall sebesar 88.82% adalah pencapaian tertinggi dalam model ini. Artinya, dari 100 kali kejadian hujan di Kota Tegal, model ANN mampu mendeteksi hampir 89 di antaranya. Dalam konteks pengembangan sistem peringatan dini banjir (Flood Early Warning System), Recall adalah metrik "hidup dan mati". Kegagalan mendeteksi hujan (False Negative) dapat berakibat fatal karena masyarakat tidak memiliki persiapan menghadapi bencana. Oleh karena itu, kemampuan model meminimalkan False Negative hingga serendah mungkin membuktikan validitas model ini untuk tujuan keselamatan publik.

2. Toleransi terhadap Presisi

Nilai Precision sebesar 76.77% menunjukkan bahwa terdapat margin kesalahan berupa "Alarm Palsu" (False Alarm). Sekitar 23% dari peringatan hujan yang dikeluarkan model mungkin tidak terbukti. Namun, dalam manajemen risiko, biaya (cost) untuk sebuah alarm palsu jauh lebih rendah dibandingkan biaya akibat bencana yang tidak terprediksi. Masyarakat yang bersiap membawa payung namun tidak jadi hujan (dampak dari presisi rendah) jauh lebih baik daripada masyarakat yang banjir tanpa peringatan (dampak dari recall rendah).

3. Keseimbangan F1-Score

Nilai F1-Score sebesar 82.35% menegaskan bahwa meskipun model memprioritaskan Recall, model tidak mengorbankan Precision secara ekstrem. Model tetap menjaga



keseimbangan yang harmonis, menjadikannya alat prediksi yang robust (tangguh) dan layak diandalkan.

Berdasarkan analisis di atas, dapat disimpulkan bahwa arsitektur ANN yang diusulkan, meskipun sederhana, sangat efektif dalam menangkap dinamika cuaca lokal Kota Tegal. Tingginya nilai Recall menjadikan model ini kandidat yang ideal untuk diintegrasikan ke dalam sistem informasi kebencanaan daerah.

Dalam konteks manajemen bencana, *False Negative* (gagal prediksi hujan) jauh lebih berbahaya daripada *False Alarm*. Efektivitas sistem peringatan dini sangat bergantung pada kemampuannya memberikan sinyal seawal mungkin, meskipun terkadang terjadi alarm palsu, agar masyarakat memiliki waktu untuk evakuasi atau persiapan [9]. Oleh karena itu, *Recall* 88.82% pada model ini dianggap sebagai indikator keberhasilan utama.

KESIMPULAN

Berdasarkan hasil pengujian, algoritma Artificial Neural Network (ANN) terbukti efektif dalam memprediksi kejadian hujan harian di Kota Tegal menggunakan data sekunder NASA POWER, dengan capaian nilai AUC sebesar 0.828. Dataset menunjukkan distribusi kelas yang seimbang secara alami, sehingga meminimalisir bias prediksi. Temuan utama penelitian ini adalah tingginya nilai Recall sebesar 88.82%, yang mengindikasikan model sangat sensitif dalam mendeteksi potensi hujan nyata. Karakteristik ini menjadikan model yang diusulkan sangat layak untuk diintegrasikan ke dalam sistem peringatan dini banjir, di mana kemampuan mendeteksi bahaya lebih diprioritaskan dibandingkan menghindari alarm palsu.

DAFTAR PUSTAKA

- F. R. Usman, W. Ridwan, and I. Z. Nasibu, "Sistem peringatan dini bencana banjir berbasis mikrokontroler Arduino," *Jambura Journal of Electrical and Electronics Engineering*, vol. 1, no. 1, pp. 1–6, 2019.
- Lisman, I. Devi Sara, and Tarmizi, "Analisis Potensi Matahari dan Angin menggunakan Reanalysis Dataset ECMWF dan NASA POWER di Pulau Teupah Lisman, Analisis," *Dielektrika – Department of Electrical Engineering University of Mataram*, vol. 10, no. 2, pp. 89–99, 2023.
- Nurmala, Y. Fitri, and S. Gautami, "Artificial Neural Network based Rainfall Prediction using Back Propagation Technique in Pekanbaru city," *PHOTON Journal of Natural Sciences and Technology*, Jan. 2024.
- N. R. Wibowo, "EFEKTIVITAS PENERAPAN SISTEM PERINGATAN DINI DALAM UPAYA PENGURANGAN RISIKO BENCANA BANJIR DI PROVINSI DKI JAKARTA," 2023.
- R. Oktafiani, A. Hermawan, and D. Avianto, "Pengaruh Komposisi Split data Terhadap Performa Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Machine Learning," *Jurnal Sains dan Informatika*, pp. 19–28, Jun. 2023, doi: 10.34128/jsi.v9i1.622.
- R. Utami and A. Hidayatullah, "The Prediksi Curah Hujan Pada Stasiun BMKG (CITEKO) Menggunakan Metode Backpropagation Neural Network," *Elkom: Jurnal Elektronika dan Komputer*, vol. 17, no. 1, pp. 251–255, 2024.
- S. Nika Purnomo, P. Katalisa, M. Purti, W. Widiyanto, and S. Sugiarto, "PENGUNAAN NASA POWER DALAM ESTIMASI DEBIT BANJIR DI DAS GARANG," 2023.



- T. Tamaji, Y. A. K. Utama, and J. Sidharta, “Jaringan Saraf Tiruan Menggunakan Metode Backpropagation untuk Prediksi Curah Hujan,” *Telekontran: Jurnal Ilmiah Telekomunikasi, Kendali dan Elektronika Terapan*, vol. 10, no. 1, pp. 30–37, 2022.
- W. O. N. R. Annisa and A. M. Sajiah, “Penerapan Jaringan Syaraf Tiruan Backpropagation pada Peramalan Curah Hujan di Kabupaten Muna,” *Jurnal Informatika Ilmu Komputer dan Sistem Informasi*, vol. 3, no. 2, 2025.