Griya Journal of Mathematics Education and Application
Volume 6 Nomor 3, September 2026
e-ISSN 2776-124X| |p-ISSN 2776-1258
https://mathjournal.unram.ac.id/index.php/Griya/index

Pemodelan prediksi curah hujan bulanan menggunakan LSTM berbasis deret waktu di Provinsi Sumatera Utara

Paskah Abadi Simanullang[1], Peter Tymoty Hutabarat[2], Rina Intan Amelia[3], Adinda Saputri[4], Regina Grace Olivia Purba[5] Suvriadi Panggabean[6]

[1,2] *Ilmu Komputer, FMIPA, Universitas Negeri Medan, Medan*
[3,4,5,6] *Matematika, FMIPA, Universitas Negeri Medan, Medan*

paskahsimanullang24@gmail.com

Diterima: 21-07-2026; Direvisi: 15-09-2026; Dipublikasi: 17-09-2026

Abstract
Rainfall prediction is important for supporting hydrometeorological disaster mitigation and data-based decision making. This study aims to model monthly rainfall prediction using Long Short-Term Memory (LSTM) based on time series data in North Sumatra Province, with monthly weather data from Medan City as the study case. The data used cover the period from 2010 to 2025 and include rainfall, rainy days, temperature, humidity, wind speed, and air pressure. The research stages consist of data validation, exploratory data analysis, threshold determination for high rainfall, feature engineering, chronological data splitting, scaling, LSTM model training, baseline comparison, backtesting, cross validation, and rainfall prediction for 2026-2027. The best LSTM configuration used 64 units, dropout 0.1, learning rate 0.0005, batch size 8, and weighted MSE loss function. Evaluation on the 2025 test set produced MAE of 75.30 mm, RMSE of 89.34 mm, MAPE of 64.23%, and R2 of 0.3703. The LSTM model outperformed the baseline models based on MAE, RMSE, and R2. Forecast results indicate higher rainfall tendencies at the end of 2026 and 2027, especially in October, November, and December.

Keywords: monthly rainfall prediction; long short-term memory; time series; feature engineering; North Sumatra

Abstrak
Prediksi curah hujan penting untuk mendukung mitigasi bencana hidrometeorologi dan pengambilan keputusan berbasis data. Penelitian ini bertujuan untuk memodelkan prediksi curah hujan bulanan menggunakan Long Short-Term Memory (LSTM) berbasis data deret waktu di Provinsi Sumatera Utara, dengan data cuaca bulanan Kota Medan sebagai studi kasus. Data yang digunakan mencakup periode 2010 hingga 2025, meliputi curah hujan, hari hujan, suhu, kelembapan, kecepatan angin, dan tekanan udara. Tahapan penelitian meliputi validasi data, exploratory data analysis, penentuan threshold curah hujan tinggi, feature engineering, pembagian data secara kronologis, scaling, pelatihan model LSTM, perbandingan baseline, backtesting, cross validation, serta prediksi curah hujan tahun 2026-2027. Konfigurasi LSTM terbaik menggunakan 64 unit, dropout 0,1, learning rate 0,0005, batch size 8, dan loss function weighted MSE. Evaluasi pada test set 2025 menghasilkan MAE 75,30 mm, RMSE 89,34 mm, MAPE 64,23%, dan R2 0,3703. Model LSTM mengungguli baseline berdasarkan MAE, RMSE, dan R2. Hasil prediksi menunjukkan kecenderungan peningkatan curah hujan pada akhir tahun 2026 dan 2027, terutama pada Oktober, November, dan Desember.

Kata Kunci: prediksi curah hujan bulanan; *long short-term memory*; deret waktu; *feature engineering*; Sumatera Utara

1. PENDAHULUAN

Curah hujan merupakan salah satu unsur iklim yang memiliki peran penting terhadap kondisi lingkungan, ketersediaan air, serta potensi terjadinya bencana hidrometeorologi seperti banjir dan tanah longsor. Curah hujan dengan intensitas tinggi dapat meningkatkan aliran permukaan dan kejenuhan tanah sehingga berpotensi memperbesar risiko bencana, terutama pada wilayah dengan kondisi geografis yang rentan (Suhadi et al., 2023). Provinsi Sumatera Utara memiliki karakteristik wilayah yang beragam, meliputi daerah pegunungan, dataran rendah, kawasan perkotaan, serta wilayah dengan banyak aliran sungai, sehingga dinamika curah hujan menjadi aspek penting untuk dikaji dalam konteks mitigasi bencana dan perencanaan wilayah (Tampubolon, 2018). Selain itu, perubahan iklim global turut menyebabkan pola curah hujan menjadi semakin tidak menentu dari segi waktu, durasi, dan intensitas, sehingga proses prediksi curah hujan menjadi lebih menantang dibandingkan sebelumnya (Tahroudi, 2025).

Kota Medan dipilih sebagai lokasi kajian karena merupakan salah satu pusat aktivitas perkotaan di Provinsi Sumatera Utara dan memiliki kebutuhan terhadap informasi curah hujan yang berkaitan dengan pengelolaan lingkungan, sumber daya air, serta mitigasi bencana hidrometeorologi. Ketersediaan data cuaca bulanan yang konsisten dalam periode 2010–2025 juga mendukung penerapan pemodelan deret waktu menggunakan Long Short-Term Memory (LSTM). Dalam penelitian ini, Kota Medan digunakan sebagai studi kasus untuk menggambarkan penerapan model prediksi curah hujan pada wilayah perkotaan di Sumatera Utara, bukan sebagai representasi penuh seluruh wilayah provinsi. Pemilihan lokasi tersebut juga memungkinkan penelitian untuk memanfaatkan data meteorologi historis sebagai dasar dalam membangun model prediksi. Penelitian sebelumnya juga telah menerapkan pendekatan pemodelan prediktif terhadap data cuaca di Kota Medan dan Kabupaten Deli Serdang, yang menunjukkan bahwa karakteristik cuaca lokal di wilayah tersebut dapat dianalisis menggunakan pendekatan berbasis data (Sianturi et al., 2025). Temuan tersebut memperkuat relevansi penggunaan data meteorologi lokal Kota Medan sebagai dasar pemodelan curah hujan dalam penelitian ini.

Kondisi curah hujan di wilayah Sumatera Utara menunjukkan adanya variasi yang cukup besar sehingga diperlukan pendekatan prediksi berbasis data. Data Badan Pusat Statistik Provinsi Sumatera Utara menunjukkan bahwa pada tahun 2024 jumlah curah hujan yang tercatat pada beberapa stasiun pengamatan BMKG berada pada kisaran 2.236–4.760 mm per tahun, dengan jumlah hari hujan mencapai 241–317 hari per tahun (Badan Pusat Statistik Provinsi Sumatera Utara, 2025). Pada Stasiun Balai Besar Meteorologi Klimatologi dan Geofisika Wilayah I di Medan, misalnya, tercatat curah hujan sebesar 3.371 mm dengan 272 hari hujan pada tahun tersebut. Tingginya jumlah curah hujan tersebut menunjukkan bahwa dinamika hujan merupakan aspek penting untuk dikaji, khususnya dalam kaitannya dengan potensi bencana hidrometeorologi.

Kondisi tersebut juga terlihat dari adanya kejadian hujan lebat yang berdampak pada banjir dan gangguan transportasi di Kota Medan (Badan Meteorologi, Klimatologi, dan Geofisika, 2024). Oleh karena itu, diperlukan model yang mampu memanfaatkan pola historis curah hujan untuk menghasilkan informasi prediksi pada periode mendatang.

Ketidakpastian pola hujan menjadi permasalahan penting karena prediksi yang kurang akurat dapat menghambat proses pengambilan keputusan, khususnya dalam perencanaan mitigasi bencana, pengelolaan sumber daya air, dan penyusunan sistem peringatan dini (Supriatna et al., 2025). Oleh karena itu, prediksi curah hujan yang akurat diperlukan untuk memberikan gambaran awal mengenai potensi kondisi hujan pada periode mendatang, sehingga pemerintah, instansi terkait, maupun masyarakat dapat melakukan langkah antisipatif secara lebih baik (Dhuha et al., 2024). Namun, data curah hujan memiliki sifat nonlinier, fluktuatif, dan dipengaruhi oleh banyak faktor atmosfer, sehingga pendekatan statistik konvensional memiliki keterbatasan dalam menangkap pola temporal yang kompleks (Rizki et al., 2020).

Penelitian prediksi curah hujan sebelumnya banyak menggunakan metode statistik deret waktu seperti ARIMA dan regresi linier, serta metode fuzzy time series (Misshuari et al., 2023; Pebralia, 2022; Prasetyo et al., 2023). Metode tersebut relatif sederhana dan dapat digunakan untuk memodelkan pola historis curah hujan. Meskipun demikian, pendekatan statistik memiliki keterbatasan dalam memodelkan hubungan nonlinier dan ketergantungan jangka panjang pada data curah hujan, terutama ketika data menunjukkan pola musiman dan fluktuasi yang tinggi (Ruslana et al., 2021). Selain metode statistik, berbagai pendekatan machine learning juga mulai digunakan untuk prediksi curah hujan karena kemampuannya dalam memodelkan hubungan yang lebih kompleks. Namun, beberapa pendekatan tersebut masih membutuhkan rekayasa fitur secara manual dan belum sepenuhnya dirancang untuk menangkap ketergantungan temporal jangka panjang pada data deret waktu (Jansen et al., 2023).

Perkembangan deep learning mendorong penggunaan Long Short-Term Memory (LSTM), yaitu pengembangan dari Recurrent Neural Network yang dirancang untuk mempelajari data berurutan dan mengatasi permasalahan hilangnya informasi pada ketergantungan jangka panjang melalui mekanisme memori dan sistem gerbang (Sari et al., 2024; Selle et al., 2022). LSTM dinilai sesuai untuk pemodelan data deret waktu karena mampu mempelajari pola temporal dan ketergantungan antarperiode. Penerapan LSTM pada pemodelan hidrologi juga menunjukkan kemampuannya dalam mempelajari hubungan temporal pada data berbasis waktu (Kratzert et al., 2018). Dalam konteks prediksi curah hujan, penelitian sebelumnya menunjukkan bahwa LSTM dapat digunakan untuk menghasilkan prediksi dengan memanfaatkan pola historis curah hujan (Jannah et al., 2022; Wijayanto et al., 2024).

Pemilihan LSTM dalam penelitian ini didasarkan pada karakteristik data yang digunakan, yaitu data curah hujan bulanan yang memiliki urutan temporal dan kemungkinan ketergantungan antarperiode. Dibandingkan pendekatan deep learning

lainnya, LSTM memiliki mekanisme memori dan gerbang yang dirancang untuk mempertahankan informasi yang relevan dari periode sebelumnya sehingga sesuai untuk memodelkan ketergantungan temporal pada data deret waktu. Arsitektur seperti GRU juga dirancang untuk menangani ketergantungan temporal dengan struktur yang lebih sederhana, sedangkan BiLSTM memanfaatkan informasi dari arah maju dan mundur dalam suatu urutan. Sementara itu, CNN-LSTM mengombinasikan ekstraksi pola lokal dengan pemodelan temporal, dan Transformer menggunakan mekanisme attention untuk mempelajari hubungan antarelemen dalam urutan. Penelitian ini memilih LSTM karena fokus utama penelitian adalah pemodelan pola temporal curah hujan bulanan dengan jumlah observasi yang tersedia, sehingga digunakan arsitektur yang secara khusus dirancang untuk data sekuensial dan memiliki struktur yang sesuai dengan kebutuhan pemodelan penelitian. Pemilihan tersebut tidak dimaksudkan untuk menyatakan bahwa LSTM selalu lebih unggul dibandingkan arsitektur deep learning lainnya, tetapi didasarkan pada kesesuaian karakteristik LSTM dengan tujuan dan karakteristik data penelitian.

Berdasarkan tinjauan tersebut, masih terdapat ruang pengembangan dalam penerapan LSTM untuk prediksi curah hujan bulanan pada wilayah Sumatera Utara, khususnya dengan memanfaatkan data cuaca bulanan historis sebagai basis pemodelan deret waktu. Penelitian ini menggunakan data cuaca bulanan Kota Medan sebagai lokasi kajian di Provinsi Sumatera Utara, dengan variabel utama berupa curah hujan bulanan dan variabel pendukung seperti hari hujan, suhu, kelembapan, tekanan udara, dan kecepatan angin. Data tersebut kemudian diolah melalui tahapan cleaning, validasi, exploratory data analysis, penentuan threshold curah hujan tinggi, feature engineering, pembentukan sequence deret waktu, pelatihan model LSTM, serta evaluasi terhadap model baseline.

Kebaruan penelitian ini terletak pada penerapan model LSTM berbasis deret waktu untuk memprediksi curah hujan bulanan dengan mengombinasikan variabel curah hujan dan variabel cuaca pendukung, serta membandingkan hasilnya dengan beberapa model baseline. Selain itu, penelitian ini juga menggunakan threshold curah hujan tinggi untuk memberikan interpretasi tambahan terhadap potensi bulan dengan curah hujan tinggi. Dengan demikian, penelitian tidak hanya berfokus pada nilai prediksi curah hujan, tetapi juga pada pemahaman pola temporal dan indikasi potensi curah hujan tinggi pada periode mendatang.

Penelitian ini memiliki beberapa batasan yang perlu diperhatikan dalam menginterpretasikan hasil. Pertama, data yang digunakan terbatas pada data cuaca bulanan Kota Medan periode Januari 2010 sampai Desember 2025, sehingga hasil penelitian belum dapat digunakan untuk menggambarkan variasi curah hujan di seluruh wilayah Provinsi Sumatera Utara. Kedua, variabel yang digunakan terutama berasal dari data meteorologi lokal, sehingga penelitian belum memasukkan variabel iklim berskala besar seperti El Niño–Southern Oscillation (ENSO), Indian Ocean Dipole (IOD),

dan Southern Oscillation Index (SOI) yang berpotensi memengaruhi pola curah hujan. Ketiga, prediksi periode 2026–2027 menggunakan estimasi variabel meteorologi pendukung berdasarkan rata-rata klimatologi bulanan historis karena data observasi masa depan belum tersedia. Oleh karena itu, hasil prediksi dalam penelitian ini diposisikan sebagai proyeksi berbasis data historis dan bukan sebagai prediksi deterministik yang menggambarkan kondisi aktual secara pasti.

Penelitian ini bertujuan untuk membangun dan mengevaluasi model prediksi curah hujan bulanan menggunakan Long Short-Term Memory (LSTM) berbasis data deret waktu di Provinsi Sumatera Utara. Secara khusus, penelitian ini bertujuan untuk menganalisis karakteristik data cuaca bulanan, membangun model prediksi curah hujan menggunakan LSTM, membandingkan kinerja model dengan baseline, serta menghasilkan prediksi curah hujan bulanan untuk periode 2026–2027. Hasil penelitian ini diharapkan dapat memberikan kontribusi terhadap pengembangan metode prediksi curah hujan berbasis kecerdasan buatan, sekaligus menjadi informasi pendukung dalam perencanaan mitigasi bencana hidrometeorologi dan pengambilan keputusan berbasis data.

2. METODE PENELITIAN

2.1 Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan pemodelan prediktif berbasis deret waktu. Tujuan penelitian adalah membangun model prediksi curah hujan bulanan menggunakan Long Short-Term Memory (LSTM). Pendekatan deret waktu digunakan karena curah hujan bulanan memiliki urutan temporal, sehingga kondisi pada suatu bulan dapat dipengaruhi oleh pola bulan-bulan sebelumnya.

Penelitian ini diarahkan untuk mendukung analisis curah hujan tinggi dalam konteks mitigasi bencana hidrometeorologi di Provinsi Sumatera Utara. Implementasi pemodelan menggunakan data cuaca bulanan Kota Medan sebagai studi kasus wilayah observasi, sehingga hasil penelitian dibahas sebagai dasar pemodelan yang dapat dikembangkan lebih lanjut pada cakupan wilayah yang lebih luas.

2.2 Data Penelitian

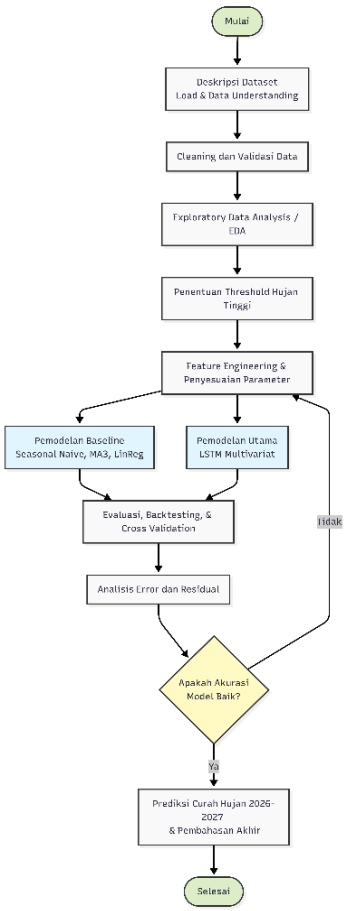
Data yang digunakan adalah data cuaca bulanan Kota Medan periode Januari 2010 sampai Desember 2025, dengan total 192 observasi. Variabel target adalah Curah_Hujan_mm, sedangkan variabel pendukung meliputi Hari_Hujan, Suhu_Rata2, Suhu_Maks_Rata2, Suhu_Min_Rata2, Kelembapan_Rata2, Angin_Rata2, dan Tekanan_Rata2.

Data divalidasi untuk memastikan kelengkapan periode, ketiadaan missing value, ketiadaan duplikasi tanggal, dan kewajaran rentang nilai meteorologi. Validasi ini

penting karena kesalahan data pada deret waktu dapat memengaruhi pembentukan fitur lag, fitur rolling, dan sequence LSTM.

2.3 Tahapan Penelitian

Tahapan penelitian disusun secara berurutan mulai dari pengumpulan data, validasi dan eksplorasi data, pembentukan fitur deret waktu, pemodelan LSTM, evaluasi model, hingga prediksi curah hujan tahun 2026-2027. Alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Flowchart Alur Penelitian

Tahap awal dilakukan dengan mengumpulkan data cuaca bulanan Kota Medan, kemudian melakukan cleaning, validasi, dan exploratory data analysis. Setelah itu, threshold curah hujan tinggi ditentukan menggunakan persentil ke-90 data training, sedangkan feature engineering dilakukan dengan membentuk fitur kalender, lag, rolling, dan feature importance.

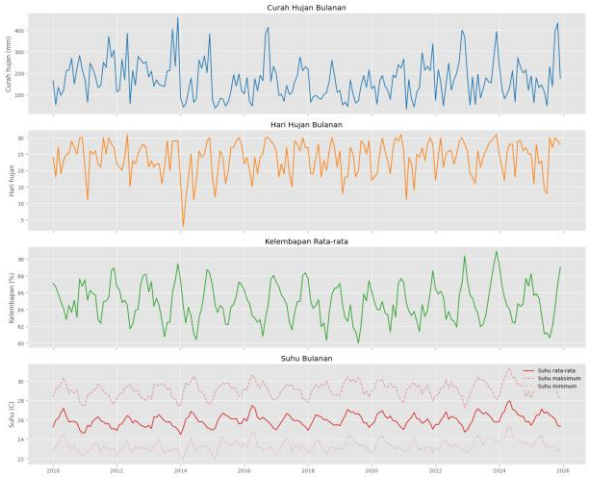
Data kemudian dibagi secara kronologis menjadi training 2010-2023, validation 2024, dan test 2025. Model baseline dan LSTM dilatih menggunakan data historis, dengan input sequence 12 bulan dan 52 fitur. Berdasarkan hasil eksperimen hyperparameter pada data validation, konfigurasi LSTM final menggunakan 64 unit, dropout 0,1, learning rate 0,0005, batch size 8, optimizer Adam, dan loss function weighted MSE.

Evaluasi dilakukan menggunakan MAE, RMSE, MAPE, R2, backtesting one-step ahead, dan time series cross validation. Setelah model final dievaluasi, model digunakan untuk memprediksi curah hujan bulanan Januari 2026 sampai Desember 2027 menggunakan metode *recursive multi-step forecasting*. Karena data meteorologi masa depan belum tersedia, variabel pendukung pada periode prediksi diestimasi menggunakan rata-rata klimatologi bulanan.

3. HASIL DAN PEMBAHASAN

3.1 Deskripsi, Validasi, dan Eksplorasi Data

Dataset terdiri dari 192 observasi bulanan periode Januari 2010 sampai Desember 2025. Variabel target adalah Curah_Hujan_mm, sedangkan variabel pendukung meliputi hari hujan, suhu, kelembapan, angin, dan tekanan udara. Setelah proses cleaning dan validasi, jumlah data tetap 192 observasi, tidak ditemukan missing value, dan setiap tahun memiliki 12 data bulanan lengkap.



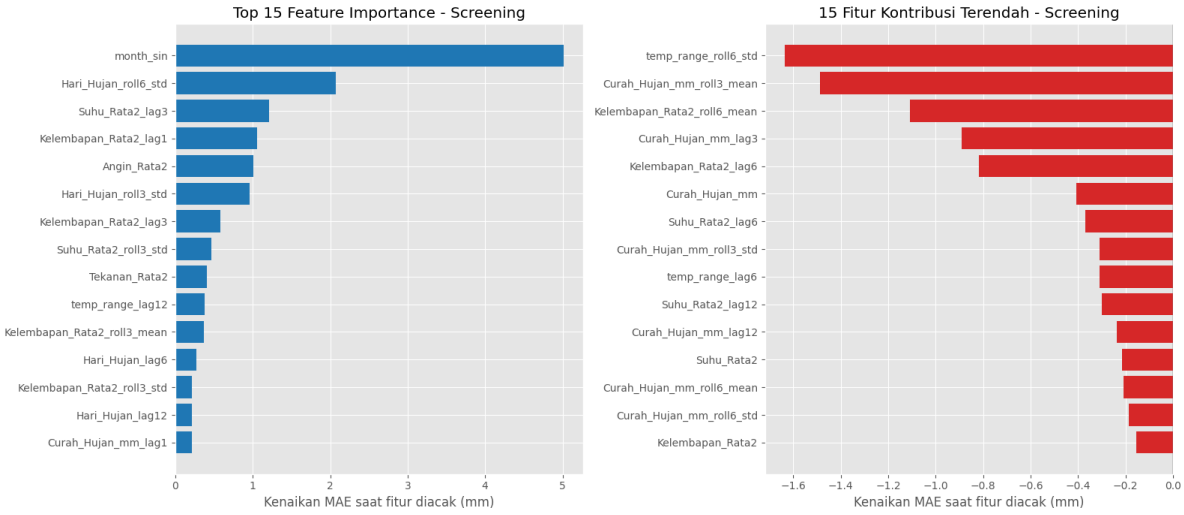
Gambar 2. Grafik Deret Waktu Curah Hujan dan Variabel Meteorologi

Gambar 2 menunjukkan bahwa curah hujan bulanan memiliki fluktuasi yang cukup besar antarbulan. Pola tersebut mendukung penggunaan pendekatan deret waktu karena prediksi curah hujan perlu mempertimbangkan informasi historis, pola musiman, dan variabel meteorologi pendukung. Korelasi awal juga menunjukkan bahwa Hari_Hujan memiliki hubungan positif paling kuat terhadap curah hujan, yaitu sekitar 0,71, sedangkan kelembapan memiliki korelasi positif sekitar 0,43.

3.2 Threshold Curah Hujan Tinggi dan Feature Engineering

Threshold curah hujan tinggi ditentukan menggunakan persentil ke-90 dari data training agar tidak terjadi data leakage. Nilai threshold yang diperoleh adalah 276,35 mm, sehingga bulan dengan curah hujan sama dengan atau lebih besar dari nilai tersebut dikategorikan sebagai bulan dengan curah hujan tinggi.

Feature engineering dilakukan dengan membentuk fitur kalender, fitur lag, dan fitur rolling. Total fitur awal yang digunakan adalah 52 fitur. Jumlah ini relatif besar dibandingkan ukuran dataset bulanan, sehingga feature importance digunakan untuk melihat fitur yang relatif berpengaruh dan fitur yang berpotensi kurang informatif.



Gambar 3. *Feature Importance* Hasil *Feature Engineering*

Berdasarkan Gambar 3, fitur musiman dan beberapa fitur berbasis hari hujan, suhu, kelembapan, dan angin memiliki kontribusi penting terhadap prediksi. Beberapa fitur dengan kontribusi rendah tidak langsung dihapus karena pengaruhnya pada model LSTM perlu diuji melalui ablation test. Dengan demikian, hasil feature importance digunakan sebagai dasar interpretasi dan rekomendasi pengembangan model, bukan sebagai keputusan penghapusan fitur pada eksperimen utama.

3.3 Pembagian Data, Baseline, dan Arsitektur LSTM

Data dibagi secara kronologis menjadi training 2010-2023, validation 2024, dan test 2025. Setelah feature engineering dan pembentukan sequence, model menggunakan lookback 12 bulan dengan 52 fitur untuk memprediksi curah hujan bulan berikutnya. Scaling dilakukan menggunakan MinMaxScaler yang hanya di-fit pada data training untuk mencegah data leakage.

Baseline yang digunakan adalah Seasonal Naive, Moving Average 3 bulan, dan Linear Regression. Model utama menggunakan LSTM final dengan satu layer LSTM 64 unit, dropout 0,1, dense layer 16 neuron, dan satu output neuron. Arsitektur ini dipilih berdasarkan hasil eksperimen hyperparameter pada data validation sehingga model yang digunakan pada tahap evaluasi sesuai dengan output notebook.

3.4 Training, Hyperparameter, dan Evaluasi Model

Eksperimen hyperparameter dilakukan secara manual dengan membandingkan jumlah unit LSTM, dropout, learning rate, dan batch size. Gambar 4 menampilkan output eksperimen sesuai notebook. Konfigurasi terbaik dipilih berdasarkan RMSE validation terendah, yaitu model LSTM_64_d0p1_lr5e-4_bs8 dengan 64 unit LSTM, dropout 0,1, learning rate 0,0005, batch size 8, dan best epoch 8.

	Model	Units_1	Dropout	LR	Batch_Size	Loss	Best_Epoch	MAE	RMSE	MAPE_%	R2
0	LSTM_64_d0p1_lr5e-4_bs8	64	0.1	0.0005	8	weighted_mse	8	50.843017	60.532561	43.541377	0.145994
1	LSTM_32_d0p1_lr1e-3_bs8	32	0.1	0.0010	8	weighted_mse	1	56.874144	62.582841	43.220083	0.087162
2	LSTM_64_d0p2_lr1e-3_bs16	64	0.2	0.0010	16	weighted_mse	3	51.723405	64.983831	48.184416	0.015777
3	LSTM_16_d0p1_lr1e-3_bs8	16	0.1	0.0010	8	weighted_mse	10	53.923836	68.814107	45.955232	-0.103667
4	LSTM_16_d0p2_lr5e-4_bs16	16	0.2	0.0005	16	weighted_mse	1	54.889575	69.696730	48.320428	-0.132160
5	LSTM_32_d0p2_lr5e-4_bs16	32	0.2	0.0005	16	weighted_mse	26	71.979598	86.686290	61.131638	-0.751394

Gambar 4. Hasil Eksperimen Hyperparameter Model LSTM

Evaluasi akhir pada test set 2025 menunjukkan bahwa LSTM memperoleh MAE 75,30 mm, RMSE 89,34 mm, MAPE 64,23%, dan R2 0,3703. Dibandingkan baseline terbaik berdasarkan MAE, yaitu Seasonal Naive, model LSTM menurunkan MAE sekitar 5,28 mm atau 6,6%.

Hasil evaluasi menunjukkan bahwa model LSTM memberikan kinerja yang lebih baik dibandingkan baseline pada metrik MAE, RMSE, dan R^2. Temuan ini menunjukkan bahwa pemanfaatan informasi temporal melalui sequence 12 bulan dapat membantu model dalam mempelajari pola curah hujan antarperiode. Hasil tersebut sejalan dengan penelitian Akbar et al. (2024) yang menerapkan LSTM untuk prediksi curah hujan bulanan pada beberapa lokasi di sekitar Medan dan menunjukkan bahwa LSTM mampu memberikan hasil prediksi yang kompetitif dibandingkan SARIMA. Penelitian Gaol et al. (2024) juga menunjukkan bahwa LSTM menghasilkan RMSE yang lebih rendah dibandingkan Multi Layer Perceptron dalam prediksi curah hujan bulanan. Selain itu, penelitian Triwahyu dan Firmansyah (2026) menunjukkan bahwa SARIMA mampu merepresentasikan pola musiman dan dinamika curah hujan bulanan di Provinsi Jawa Barat. Temuan tersebut memperkuat bahwa pola musiman merupakan komponen penting dalam pemodelan curah hujan bulanan, baik pada pendekatan statistik maupun deep learning. Dengan demikian, hasil penelitian ini memperkuat temuan bahwa

arsitektur yang mempertimbangkan urutan temporal dapat digunakan untuk memodelkan curah hujan bulanan.

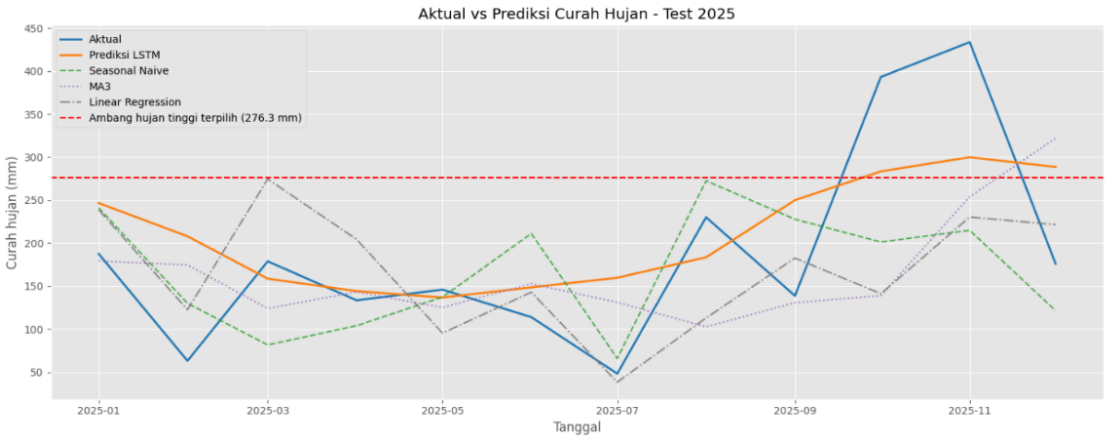
Meskipun demikian, peningkatan kinerja LSTM pada penelitian ini relatif moderat, yaitu penurunan MAE sebesar 5,28 mm atau sekitar 6,6% dibandingkan Seasonal Naive. Hasil tersebut menunjukkan bahwa pola musiman pada curah hujan Kota Medan masih cukup kuat sehingga baseline sederhana yang memanfaatkan pola historis musiman tetap mampu memberikan prediksi yang kompetitif. Temuan ini juga menunjukkan bahwa penggunaan LSTM tidak secara otomatis menghasilkan peningkatan kinerja yang sangat besar dibandingkan metode sederhana. Hal tersebut sejalan dengan penelitian Aslam dan Afghani (2024) di Sumatera Barat yang menunjukkan bahwa LSTM tidak selalu menjadi model dengan kinerja terbaik ketika dibandingkan dengan metode statistik dan machine learning lainnya. Dalam penelitian tersebut, XGBoost menghasilkan kinerja yang lebih baik dibandingkan SARIMA, ETS, dan LSTM. Oleh karena itu, hasil penelitian ini mendukung pentingnya membandingkan LSTM dengan baseline sebelum menyimpulkan efektivitas model.

	Model	MAE	RMSE	MAPE_%	R2
0	LSTM Final	75.304058	89.343539	64.227199	0.370255
1	Seasonal Naive	80.575833	101.969031	45.958250	0.179696
2	MA3	86.794444	115.023993	57.403590	-0.043795
3	Linear Regression	85.658662	110.399165	43.966911	0.038454

Gambar 5. Perbandingan Metrik LSTM dan Baseline pada Test Set 2025

Hasil tersebut menunjukkan bahwa LSTM unggul pada MAE, RMSE, dan R2, tetapi belum unggul pada MAPE. Nilai MAPE yang tinggi perlu ditafsirkan hati-hati karena curah hujan bulanan dapat bernilai rendah pada beberapa bulan, sehingga selisih prediksi dalam milimeter dapat terlihat besar dalam bentuk persentase.

Nilai MAPE sebesar 64,23% menunjukkan bahwa kesalahan relatif model masih cukup tinggi apabila dievaluasi dalam bentuk persentase. Namun, interpretasi metrik tersebut perlu mempertimbangkan karakteristik data curah hujan yang memiliki variasi nilai cukup besar antarbulan. Pada bulan dengan curah hujan aktual yang rendah, selisih prediksi yang secara absolut tidak terlalu besar dapat menghasilkan nilai kesalahan persentase yang tinggi. Oleh karena itu, MAPE dalam penelitian ini tidak digunakan sebagai satu-satunya dasar untuk menentukan model terbaik, tetapi diinterpretasikan bersama MAE, RMSE, dan R^2. Pendekatan tersebut penting karena MAE dan RMSE memberikan informasi kesalahan dalam satuan milimeter sehingga lebih mudah dikaitkan dengan besarnya penyimpangan prediksi curah hujan secara aktual.



Gambar 6. Grafik Aktual vs Prediksi Curah Hujan pada Test Set 2025

Grafik backtesting menunjukkan bahwa model mampu mengikuti pola umum curah hujan, terutama peningkatan pada akhir tahun. Namun, prediksi LSTM masih cenderung lebih halus dibanding nilai aktual, sehingga puncak curah hujan ekstrem belum sepenuhnya tertangkap. Analisis error menunjukkan bahwa kesalahan terbesar terjadi pada Februari, sedangkan error terkecil terjadi pada Mei.

Hasil backtesting menunjukkan bahwa prediksi LSTM cenderung lebih halus dibandingkan nilai aktual, terutama ketika terjadi peningkatan curah hujan yang tajam. Kondisi tersebut menunjukkan bahwa model lebih mampu menangkap pola umum dan kecenderungan musiman dibandingkan perubahan ekstrem dalam waktu singkat. Temuan serupa juga dilaporkan dalam penelitian Aslam dan Afghani (2024), yang menunjukkan bahwa model prediksi curah hujan masih mengalami kesulitan dalam menangkap nilai pencilan yang berkaitan dengan kondisi tidak normal dan fenomena iklim seperti ENSO dan IOD. Hasil penelitian Rafhida dan Najib (2026) di Jakarta Pusat juga menunjukkan bahwa model LSTM masih menghadapi kesulitan dalam memprediksi puncak curah hujan ekstrem (Rafhida & Najib, 2026). Dengan demikian, kemampuan LSTM dalam menangkap pola temporal belum selalu diikuti oleh kemampuan yang sama dalam memprediksi kejadian ekstrem. Kondisi ini menjadi salah satu alasan penting untuk mempertimbangkan penambahan variabel iklim berskala besar dan pendekatan pemodelan yang lebih khusus terhadap kejadian ekstrem pada penelitian selanjutnya.

	Fold	Train_End	Validation_Period	Train_Samples	Validation_Samples	Epochs_Run	MAE	RMSE	MAPE_%
0	1	2018-12	2019-01 s.d. 2020-12	84	24	12	46.659594	53.130732	50.277531
1	2	2020-12	2021-01 s.d. 2022-12	108	24	9	80.983443	96.394191	92.579980
2	3	2022-12	2023-01 s.d. 2024-12	132	24	13	63.524601	75.906834	51.124187

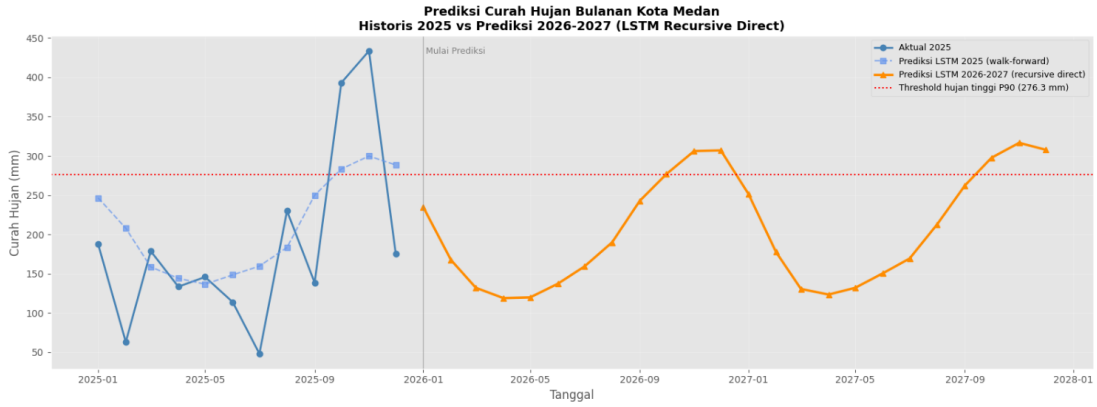
Gambar 7. Hasil Time Series Cross Validation dengan Expanding Window

Hasil time series cross validation menunjukkan variasi performa antarfold. Rata-rata MAE sebesar 70,60 mm dengan standar deviasi 26,14 mm menunjukkan bahwa performa model belum sepenuhnya stabil pada seluruh periode historis. Oleh karena itu, hasil test 2025 perlu dibaca bersama hasil cross validation.

Variasi performa antarfold pada time series cross validation menunjukkan bahwa kemampuan model tidak sepenuhnya konsisten pada seluruh periode historis. Nilai MAE rata-rata sebesar 70,60 mm dengan standar deviasi 26,14 mm menunjukkan adanya perubahan tingkat kesulitan prediksi antarperiode. Hal ini dapat terjadi karena karakteristik curah hujan tidak selalu sama pada setiap periode, terutama ketika terdapat bulan dengan peningkatan atau penurunan curah hujan yang lebih besar dibandingkan pola umumnya. Oleh karena itu, penggunaan satu periode pengujian saja berpotensi memberikan gambaran kinerja yang terlalu spesifik terhadap kondisi pada periode tersebut. Hasil cross validation dalam penelitian ini memberikan informasi tambahan bahwa kinerja model perlu dinilai pada beberapa periode waktu agar kemampuan generalisasi model dapat diamati secara lebih komprehensif.

3.5 Prediksi Curah Hujan 2026-2027 dan Pembahasan Akhir

Model final digunakan untuk memprediksi curah hujan bulanan Januari 2026 sampai Desember 2027 menggunakan metode *recursive multi-step forecasting*. Pada metode ini, prediksi dilakukan secara bertahap untuk setiap bulan, kemudian hasil prediksi bulan sebelumnya digunakan sebagai bagian dari riwayat untuk memprediksi bulan berikutnya. Karena variabel meteorologi masa depan belum tersedia, variabel pendukung pada periode prediksi diestimasi menggunakan rata-rata klimatologi bulanan historis. Oleh sebab itu, hasil prediksi dipahami sebagai proyeksi berbasis asumsi, bukan observasi aktual.



Gambar 8. Grafik Prediksi Curah Hujan Bulanan Kota Medan Tahun 2026-2027

Prediksi menunjukkan pola peningkatan curah hujan pada akhir tahun. Bulan Oktober, November, dan Desember pada tahun 2026 maupun 2027 diprediksi berada pada kisaran tinggi dan beberapa di antaranya melewati threshold P90 sebesar 276,35 mm. Pola ini konsisten dengan hasil feature importance yang menunjukkan bahwa komponen musiman berperan besar terhadap prediksi.

Pola tersebut menunjukkan bahwa model mempertahankan informasi musiman yang dipelajari dari data historis ke dalam periode prediksi. Secara metodologis, hasil ini mendukung penggunaan sequence temporal dan fitur kalender dalam model karena komponen musiman menjadi salah satu fitur yang memiliki kontribusi terhadap prediksi. Dengan demikian, prediksi 2026–2027 lebih tepat dipahami sebagai proyeksi berdasarkan pola historis daripada prakiraan operasional yang menggambarkan kondisi atmosfer aktual pada masa mendatang.

Tahun	Rata-rata (mm)	Minimum (mm)	Maksimum (mm)	Total (mm)
2026	199.4	119.0	307.0	2392.6
2027	211.1	123.5	316.6	2532.8

Gambar 9. Ringkasan Prediksi Curah Hujan Tahun 2026-2027

Keterbatasan penelitian ini juga berkaitan dengan penggunaan data Kota Medan sebagai studi kasus dalam konteks Sumatera Utara serta belum digunakannya variabel iklim skala besar seperti ENSO, IOD, atau SOI. Penelitian lanjutan disarankan untuk menambahkan variabel iklim tersebut dan melakukan *ablation test* terhadap fitur dengan kontribusi rendah.

3.6 Implikasi Teoritis dan Praktis

Secara teoretis, hasil penelitian ini menunjukkan bahwa LSTM dapat digunakan untuk memodelkan curah hujan bulanan sebagai data deret waktu dengan memanfaatkan hubungan temporal antarperiode dan variabel meteorologi pendukung. Meskipun demikian, model masih memiliki keterbatasan dalam merepresentasikan kejadian curah hujan ekstrem. Dengan demikian, penelitian ini memberikan gambaran bahwa efektivitas LSTM pada prediksi curah hujan tidak hanya ditentukan oleh arsitektur model, tetapi juga oleh karakteristik data, pemilihan fitur, pola musiman, serta keberadaan faktor iklim yang memengaruhi variabilitas curah hujan.

Secara praktis, model yang dikembangkan dapat digunakan sebagai informasi pendukung untuk memahami kecenderungan curah hujan bulanan dan mengidentifikasi periode yang berpotensi memiliki curah hujan tinggi. Informasi tersebut dapat dimanfaatkan sebagai salah satu bahan pertimbangan awal dalam perencanaan

pengelolaan sumber daya air, kesiapsiagaan terhadap bencana hidrometeorologi, dan pengembangan sistem informasi prediksi berbasis data. Namun, karena hasil prediksi 2026–2027 masih bergantung pada estimasi variabel meteorologi berdasarkan klimatologi historis dan performa model belum sepenuhnya stabil, hasil penelitian ini belum dapat digunakan sebagai satu-satunya dasar dalam pengambilan keputusan operasional atau sistem peringatan dini. Penggunaan model secara praktis perlu dikombinasikan dengan data observasi terbaru dan informasi prakiraan meteorologi dari lembaga terkait.

4. SIMPULAN

Penelitian ini menunjukkan bahwa model Long Short-Term Memory (LSTM) dapat digunakan untuk memodelkan dan memprediksi curah hujan bulanan berbasis deret waktu pada data cuaca bulanan Kota Medan sebagai studi kasus di Provinsi Sumatera Utara. Tahapan pemodelan yang mencakup validasi data, pembentukan fitur deret waktu, pembagian data secara kronologis, scaling, pelatihan model, serta evaluasi terhadap baseline mampu menghasilkan model yang dapat menangkap pola temporal curah hujan bulanan. Konfigurasi terbaik diperoleh pada model LSTM dengan 64 unit, dropout 0,1, learning rate 0,0005, batch size 8, dan loss function weighted MSE.

Hasil evaluasi pada test set tahun 2025 menunjukkan bahwa model LSTM memiliki kinerja lebih baik dibandingkan baseline berdasarkan MAE, RMSE, dan R2, meskipun nilai MAPE masih relatif tinggi karena adanya fluktuasi curah hujan bulanan dan nilai aktual yang rendah pada beberapa periode. Prediksi tahun 2026-2027 menunjukkan kecenderungan peningkatan curah hujan pada akhir tahun, terutama pada bulan Oktober, November, dan Desember, sehingga hasil model dapat menjadi informasi pendukung awal dalam memahami pola curah hujan bulanan dan potensi periode curah hujan tinggi. Namun, hasil penelitian ini tetap perlu dipahami sebagai proyeksi berbasis data historis dan asumsi klimatologis, sehingga pengembangan lebih lanjut diperlukan dengan cakupan wilayah yang lebih luas serta penambahan variabel iklim berskala besar untuk meningkatkan stabilitas dan ketepatan prediksi.

5. UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada dosen pembimbing atas bimbingan dan arahan selama pelaksanaan penelitian. Terima kasih juga disampaikan kepada rekan-rekan peneliti atas kontribusi dan kerja sama yang diberikan. Penulis turut mengucapkan terima kasih kepada Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas Negeri Medan (UNIMED) atas dukungan pendanaan melalui skema Penelitian Student Grant yang didanai oleh PNBP, sehingga penelitian ini dapat terlaksana dengan baik.

6. REKOMENDASI

Berdasarkan hasil penelitian, pengembangan selanjutnya disarankan menggunakan cakupan data yang lebih luas, tidak hanya terbatas pada Kota Medan, tetapi juga mencakup beberapa wilayah lain di Provinsi Sumatera Utara agar model dapat merepresentasikan variasi spasial curah hujan dengan lebih baik. Penelitian lanjutan juga dapat menambahkan variabel iklim berskala besar seperti ENSO, IOD, SOI, arah angin, dan indeks monsun untuk meningkatkan kemampuan model dalam menangkap faktor-faktor atmosfer yang memengaruhi curah hujan. Selain itu, perlu dilakukan ablation test terhadap fitur hasil feature engineering untuk mengetahui fitur mana yang benar-benar memberikan kontribusi signifikan terhadap kinerja model.

Dari sisi pemodelan, penelitian berikutnya dapat membandingkan LSTM dengan arsitektur deep learning lain seperti GRU, BiLSTM, CNN-LSTM, atau Transformer berbasis deret waktu. Evaluasi juga sebaiknya menggunakan periode data yang lebih panjang dan validasi lintas wilayah agar stabilitas model dapat diuji secara lebih menyeluruh. Hambatan utama dalam penelitian ini adalah keterbatasan data historis, belum tersedianya variabel meteorologi masa depan, serta ketidakpastian prediksi pada bulan dengan fluktuasi curah hujan tinggi. Oleh karena itu, hasil prediksi sebaiknya digunakan sebagai informasi pendukung awal dan bukan sebagai satu-satunya dasar pengambilan keputusan mitigasi.

7. REFERENSI

Akbar, A. A., Darmawan, Y., Wibowo, A., & Rahmat, H. K. (2024). Accuracy assessment of monthly rainfall predictions using seasonal ARIMA and long short-term memory (LSTM). *Journal of Computer Science and Engineering (JCSE), 5*(2), 99–114. https://doi.org/10.36596/jcse.v5i2.843

Aslam, F. M., & Afghani, F. A. (2024). Comparing monthly rainfall prediction in West Sumatra using SARIMA, ETS, LSTM, and XGBoosting methods. *Indonesian Journal of Applied Statistics, 7*(1), 14–26. https://doi.org/10.13057/ijas.v7i1.83187

Badan Meteorologi, Klimatologi, dan Geofisika. (2024). *Hujan lebat menyebabkan banjir dan gangguan transportasi di Kota Medan*. BMKG.

Badan Pusat Statistik Provinsi Sumatera Utara. (2025). *Jumlah curah hujan dan hari hujan menurut stasiun BMKG di Provinsi Sumatera Utara*. BPS Provinsi Sumatera Utara.

Dhuha, M. A. S., Saputra, A. H., Kristianto, A., & Mulya, A. (2024). Impact-Based Forecasting (IBF) untuk mendukung manajemen risiko banjir di kawasan Jabodetabek. *Jurnal Ilmu Lingkungan, 22*(5), 1159–1173. https://doi.org/10.14710/jil.22.5.1159-1173

Gaol, G. M. L., Syafrullah, M., & Supardi, S. (2024). Comparison of monthly rainfall prediction using long short term memory and multi layer perceptron methods in South Tangerang City. *Jurnal Sisfokom (Sistem Informasi dan Komputer), 13*(2), 293–297. https://doi.org/10.32736/sisfokom.v13i2.2149

Jannah, A. N., Fauji, A., & Wiguna, M. S. (2022). Perbandingan model statistika dan machine learning dalam memprakirakan curah hujan dasarian. *Jurnal Meteorologi dan Geofisika, 23*(2), 75–86.

Jansen, D., Handhayani, T., & Hendryli, J. (2023). Penerapan metode *long short-term memory* dalam memprediksi data meteorologi di Kalimantan Timur. *Simtek: Jurnal Sistem Informasi dan Teknik Komputer, 8*(2), 348–352. https://doi.org/10.51876/simtek.v8i2.202

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences, 22*(11), 6005–6022. https://doi.org/10.5194/hess-22-6005-2018

Misshuari, I. W., Kurniyaningrum, E., & Saily, R. (2023). Studi peramalan curah hujan menggunakan metode ARIMA di Daerah Asahan. *Construction Engineering and Sustainable Development, 6*(2), 22–28. https://doi.org/10.25105/cesd.v6i2.18815

Nazeri Tahroudi, M. (2025). Comprehensive global assessment of precipitation trend and pattern variability considering their distribution dynamics. *Scientific Reports, 15*(1), Article 22458. https://doi.org/10.1038/s41598-025-06050-5

Pebralia, J. (2022). Analisis curah hujan menggunakan machine learning metode regresi linier berganda berbasis Python dan Jupyter Notebook. *Jurnal Ilmu Fisika dan Pembelajarannya, 6*(2), 23–30. https://doi.org/10.19109/jifp.v6i2.13958

Prasetyo, H. R., Palupi, I., & Wahyudi, B. A. (2023). Prediksi menggunakan model *fuzzy time series*: Studi kasus curah hujan di Kabupaten Bandung. *LOGIC: Jurnal Penelitian Informatika, 1*(1), 8–13. https://doi.org/10.25124/logic.v1i1.6405

Rafhida, S. A., & Najib, M. K. (2026). Pemodelan curah hujan bulanan Jakarta Pusat menggunakan neural network long short-term memory (LSTM). *Jurnal Sains Matematika dan Statistika, 12*(2). https://doi.org/10.24014/t1b6xb48

Rizki, M., Basuki, S., & Azhar, Y. (2020). Implementasi deep learning menggunakan arsitektur long short-term memory (LSTM) untuk prediksi curah hujan Kota Malang. *Jurnal Repositor, 2*(3), 331–338. https://doi.org/10.22219/repositor.v2i3.470

Ruslana, Z. N., Tresnawati, R., Rosyidah, R., Harmoko, I. W., & Siswanto, S. (2021). Reliabilitas prediksi curah hujan dasarian pada kejadian curah hujan ekstrim pemicu banjir 26 Oktober 2020 di Kebumen: Model statistik (HyBMG) versus model dinamik (ECMWF). *Jurnal Geosains dan Teknologi, 4*(2), 83–100. https://doi.org/10.14710/jgt.4.2.2021.83-100

Sari, J. R., Hairani, H., & Latif, K. A. (2024). Perbandingan model gated recurrent unit dan long short-term memory untuk prediksi curah hujan. *Indonesian Journal of Computer Science and Research, 2*(2), 111–123. https://doi.org/10.59095/ijcsr.v2i2.77

Sianturi, A. H., Panjaitan, E. J. B., Kaban, F. A. B., Panggabean, I. F. J., Johana, N. L., Bangun, S. J., & Panggabean, S. (2025). Analisis probabilistik untuk prediksi cuaca di Kota Medan dan Kabupaten Deli Serdang (menggunakan metode Rantai Markov). *Griya Journal of Mathematics Education and Application, 5*(4). https://doi.org/10.29303/griya.v5i4.970

Selle, N., Yudistira, N., & Dewi, C. (2022). Perbandingan prediksi penggunaan listrik dengan menggunakan metode *long short-term memory* (LSTM) dan *recurrent neural network* (RNN). *Jurnal Teknologi Informasi dan Ilmu Komputer, 9*(1), 155–162. https://doi.org/10.25126/jtiik.202295585

Suhadi, S., Mabruroh, F., Wiyanto, A., & Ikra, I. (2023). Analisis fenomena perubahan iklim terhadap curah hujan ekstrim. *OPTIKA: Jurnal Pendidikan Fisika, 7*(1), 94–100. https://doi.org/10.37478/optika.v7i1.2738

Supriatna, D., Anggai, S., & Tukiyat, T. (2025). Analisis prediksi curah hujan di Kota Tangerang menggunakan metode LSTM dan GRU. *Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence), 5*(2), 119–131. https://doi.org/10.55382/jurnalpustakaai.v5i2.1068

Tampubolon, K. (2018). Aplikasi sistem informasi geografis (SIG) sebagai penentuan kawasan rawan banjir di Kota Medan. *Jurnal Pembangunan Perkotaan, 6*(2), 63–68. http://ejpp.balitbang.pemkomedan.go.id/index.php/JPP

Triwahyu, M., & Firmansyah, M. R. (2026). Pemodelan SARIMA terhadap curah hujan bulanan di Provinsi Jawa Barat. *Griya Journal of Mathematics Education and Application, 6*(1), 272–283. https://doi.org/10.29303/griya.v6i1.1017

Wijayanto, A., Sugiharto, A., & Santoso, R. (2024). Identifikasi dini curah hujan berpotensi banjir menggunakan algoritma long short-term memory (LSTM) dan Isolation Forest. *Jurnal Teknologi Informasi dan Ilmu Komputer, 11*(3), 637–646. https://doi.org/10.25126/jtiik.938718