

# Prediksi Curah Hujan Harian Menggunakan Metode Principal Component Analysis (PCA) dan Long Short-Term Memory (LSTM)

M. Risky Praselia<sup>1</sup>, Ratna Cintya Dewi<sup>2</sup>, Ahmad Hanif Asyhar<sup>1</sup>, Abdulloh Hamid<sup>1</sup>

<sup>1</sup>Program Studi Matematika, Universitas Islam Negeri Sunan Ampel Surabaya, Indonesia.

<sup>2</sup>Badan Meteorologi Klimatologi dan Geofisika, Indonesia.

---

## Artikel Info

### Kata Kunci:

Curah Hujan;  
Long Short-Term Memory;  
Prediksi;  
Principal Component Analysis.

### Keywords:

Rainfall;  
Principal Component Analysis;  
Prediction;  
Long Short-Term Memory.

---

### Riwayat Artikel:

Submitted: 16 Oktober 2025

Accepted: 21 Januari 2026

Published: 22 Januari 2026

**Abstrak:** Intensitas curah hujan memiliki keterkaitan erat dengan berbagai aspek kehidupan, khususnya dalam sektor lingkungan, transportasi, dan mitigasi bencana. Pengukuran curah hujan di Indonesia umumnya dilakukan oleh Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) menggunakan alat konvensional yang berpotensi mengalami penurunan akurasi akibat faktor teknis maupun lingkungan. Oleh karena itu, diperlukan metode prediksi sebagai alternatif untuk menghindari kekosongan data serta sebagai acuan pengambilan keputusan. Penelitian ini bertujuan memprediksi curah hujan harian di Tanjung Perak, Surabaya, menggunakan kombinasi Principal Component Analysis (PCA) dan Long Short-Term Memory (LSTM). PCA digunakan untuk mereduksi dimensi variabel meteorologi sehingga diperoleh fitur yang paling berpengaruh terhadap curah hujan, sedangkan LSTM digunakan untuk memodelkan data deret waktu. Data yang digunakan merupakan data observasi harian BMKG periode 1 Januari 2020 hingga 31 Desember 2024. Hasil PCA menunjukkan bahwa temperatur rata-rata, kelembapan rata-rata, dan kecepatan angin rata-rata merupakan variabel dominan. Model terbaik diperoleh pada konfigurasi 150 hidden layer, batch size 256, dan learn rate drop period 50 dengan nilai MAAPE sebesar 0,5662. Hasil tersebut menunjukkan bahwa kombinasi metode PCA-LSTM mampu memprediksi curah hujan harian dengan kinerja yang cukup baik.

**Abstract:** Rainfall intensity has a close relationship with various aspects of life, particularly in environmental management, transportation, and disaster mitigation. In Indonesia, rainfall measurements are generally conducted by the Meteorology, Climatology, and Geophysics Agency (BMKG) using conventional instruments that may experience a decline in accuracy due to technical and environmental factors. Therefore, a prediction method is required as an alternative to prevent data gaps and to serve as a reference for decision-making. This study aims to predict daily rainfall in Tanjung Perak, Surabaya, using a combination of Principal Component Analysis (PCA) and Long Short-Term Memory (LSTM). PCA is applied to reduce the dimensionality of meteorological variables and to identify the most influential features, while LSTM is used to model time series data. The dataset consists of daily observational data from BMKG covering the period from January 1, 2020, to December 31, 2024. The PCA results indicate that average temperature, average humidity, and average wind speed are the dominant variables influencing rainfall. The best model performance was achieved using 150 hidden layers, a batch size of 256, and a learn rate drop period of 50, resulting in a MAAPE value of 0.5662. These results demonstrate that the PCA-LSTM combination is capable of predicting daily rainfall with satisfactory performance.

**Corresponding Author:**

M. Risky Prasetya

Email: mr.prasetya27@gmail.com

---

## PENDAHULUAN

Cuaca merupakan suatu indikator perubahan iklim bumi yang sekaligus menjadi faktor penting dalam segala aspek kehidupan (Sunarmi et al., 2022). Disebutkan dalam *World Climate Conference* kondisi langit yang ditakar secara kompleks yang terdiri dari perubahan, perkembangan, dan hilang atau munculnya sebuah fenomena di langit disebut cuaca udara (Luthfiarta et al., 2020). Pada negara yang terletak di daerah khatulistiwa seperti Indonesia umumnya beriklim tropis dan hanya terdapat musim hujan dan tak hujan atau kemarau. Dimana besarnya intensitas curah hujan memiliki pengaruh terhadap kondisi cuaca (Kosasih et al., 2021). Hal tersebut juga dapat menjadi faktor yang penting dalam mempengaruhi terjadinya perubahan iklim.

Curah hujan merupakan sebagian dari beberapa parameter yang sangat berpengaruh dalam beberapa aspek yang ada dalam kehidupan manusia setiap harinya. Beberapa badan di Indonesia telah melakukan percobaan pengukuran terhadap intensitas curah hujan, termasuk oleh Badan Meteorologi Klimatologi dan Geofisika (BMKG) yang menggunakan alat konvensional yang dilakukan secara langsung di lapangan (Peranginangin et al., 2021). Dimana seiring berjalannya waktu, sebuah alat akan mengalami penurunan akurasi yang juga disebabkan oleh berbagai faktor. Selain menangani alat dengan melakukan kalibrasi, perlu juga dilakukan prediksi yang mencegah terjadinya kekosongan data dan bisa juga sebagai acuan dalam pengambilan sebuah keputusan terkait data curah hujan atau variabel lain (Firdaus et al., 2020). Temperatur, kelembaban, dan kecepatan angin merupakan 3 faktor utama yang berpengaruh terhadap curah hujan (Salehe et al., 2024).

Faktor yang mempengaruhi curah hujan selalu berubah setiap waktu dan relatif bersifat kompleks, sehingga diperlukan penggunaan metode yang tepat untuk menangani kasus ini. Penggunaan berbagai macam metode telah dilakukan banyak peneliti terdahulu. Penelitian yang menerapkan metode *backpropagation* guna melakukan prediksi curah hujan menghasilkan nilai MSE sebesar 0.006952 (Herdhyanti et al., 2022). Pemanfaatan metode *decision tree* untuk melakukan prediksi curah hujan dalam kasus lainnya mendapatkan akurasi sebesar 89.4% (Hasanah et al., 2021). Algoritma lain seperti *levenberg marquardt* digunakan untuk memprediksi curah hujan di Jakarta yang mendapatkan nilai akurasi 96% (Merdekawati & Ismail, 2019). Penggunaan metode lain seperti *categorical boosting* yang memperoleh nilai akurasi sebesar 94% (Istianto et al., 2024). Prediksi intensitas curah hujan dengan metode *Long Short-Term Memory* (LSTM) yang mendapatkan hasil RMSE sebesar 0.98162 dan nilai MAE sebesar 0.68847 (Ramadhan & Fachrie, 2024).

Penelitian menggunakan metode LSTM telah banyak dilakukan menggunakan berbagai macam objek yang berbeda. Penelitian yang memanfaatkan metode LSTM dalam melakukan prediksi terkait harga *cryptocurrency* menunjukkan performa terbaik dengan nilai RMSE sebesar 0.063% menggunakan *epoch* berjumlah 20 (Rizkilloh & Widiyanesti, 2022). Penggunaan metode LSTM untuk melakukan prediksi tingkat temperatur kota Semarang yang mendapatkan hasil MAPE sebesar 1.896% dan RMSE sebesar 0.725 (Akbar et al., 2023). Prediksi harga saham Bank Rakyat Indonesia dengan metode LSTM memperoleh nilai RMSE yang cukup baik yaitu 227.470 (Satyo et al., 2020). Terdapat juga penelitian tentang prediksi temperatur cuaca di negara Norwegia yang mendapatkan hasil nilai RMSE terbaiknya pada 1.242 (Hidayatullah & Cherid, 2023).

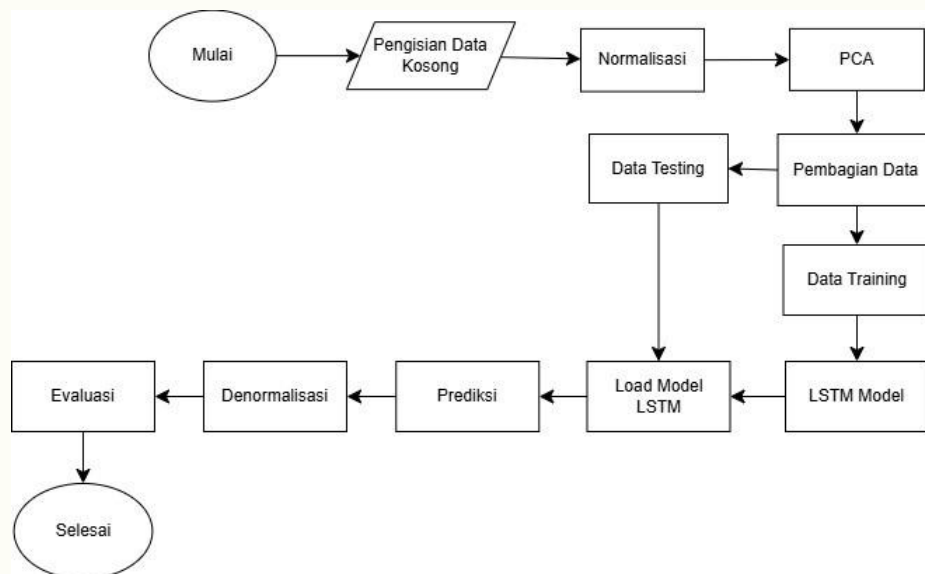
LSTM merupakan metode yang sudah terbukti berhasil dalam menangani prediksi data time series yang merupakan bentuk inovasi dari *recurrent neural network* (RNN) (Ashari & Sadikin, 2020). LSTM mampu menangani hubungan *nonlinear* terlalu tinggi seperti pada kasus curah hujan. Namun, karena terdapat beberapa faktor yang dapat mempengaruhi intensitas curah hujan dan bersifat kompleks, sehingga proses seleksi fitur diperlukan guna menentukan faktor mana saja yang paling berpengaruh. *Principal Component Analysis* (PCA) merupakan ekstraksi fitur yang meringkas atribut yang awalnya banyak menjadi lebih sedikit dengan mempertimbangkan pengaruhnya (Kusnanti et al.,

2024). Penelitian yang mencoba menerapkan PCA-LSTM untuk menentukan komponen utama dalam melakukan prediksi produktivitas kapal keruk hisap pemotong yang memperoleh nilai MAE sebesar 0.0424, nilai R-square 0.999 dan nilai RMSE 0.0925 (Yang et al., 2021). Penelitian lain tentang peramalan beban jangka pendek menggunakan PCA-LSTM menghasilkan koefisien korelasi sebesar 94% (Fang et al., 2020). Terdapat juga penelitian tentang prediksi curah hujan harian menggunakan PCA-LSTM menghasilkan nilai MAPE 0.0018 menggunakan 100 *hidden layer*, 32 *batch size*, dan 50 *learn rate* (Musfiroh et al., 2023).

Dengan kemampuan PCA yang dapat memilih fitur yang paling berpengaruh dan LSTM yang dapat mengatasi data *time series* (Amatullah et al., 2022). Berdasarkan penjelasan latar belakang tersebut, prediksi curah hujan harian pada penelitian ini akan memanfaatkan metode PCA-LSTM. Perbedaan utama penelitian ini terletak pada lokasi penelitian yaitu Tanjung Perak, Surabaya, serta komposisi variabel input yang berbeda dari penelitian terdahulu. Variabel yang digunakan meliputi kecepatan angin rata-rata, kecepatan angin maksimum, temperatur rata-rata, temperatur minimal dan maksimal, kelembaban rata-rata, dan lamanya matahari bersinar. Seluruh variabel tersebut terlebih dahulu melalui proses reduksi fitur menggunakan *Principal Component Analysis* (PCA). Variabel hasilnya digunakan sebagai input dalam model LSTM untuk melakukan prediksi curah hujan. Pemodelan dilakukan menggunakan *software Python*, *Mean Arctangent Absolute Percentage Error* (MAAPE) digunakan untuk mengevaluasi hasil prediksi.

## METODE

Riset ini dimulai dengan melakukan pengumpulan data. Dilanjutkan dengan melakukan pengisian data kosong dan melakukan normalisasi, serta pembentukan pola dari *time series*. Melakukan seleksi fitur untuk mengetahui variabel apa saja yang berpengaruh besar untuk curah hujan dengan menerapkan PCA. Kemudian dilakukan pembagian data training. Data training berguna untuk melatih dan memperoleh model dari LSTM dan selanjutnya data testing akan digunakan untuk memprediksi. Proses denormalisasi juga dilakukan guna mengembalikan nilai yang diperoleh. Evaluasi model menggunakan MAAPE. Gambar 1 merupakan penjelasan proses yang dilakukan dalam penelitian ini.



Gambar 1. Diagram Alir Penelitian

## Dataset

Data meteorologi diperoleh dari BMKG Tanjung Perak Surabaya yang didapatkan dari hasil observasi harian yang berawal 1 Januari 2020 hingga 31 Desember 2024 dengan jumlah data sebanyak 1.830 data yang dibagi menjadi 1.464 data *training* serta 366 data testing. Terdapat beberapa variabel input dalam data, yaitu temperatur minimal ( $T_n$ ), temperatur maksimal ( $T_x$ ), temperatur rata-rata

(Tavg), kelembaban rata- rata (RH\_avg), lamanya matahari bersinar (SS), kecepatan angin maksimal (FF\_x), kecepatan angin rata- rata (FF\_avg), serta variabel curah hujan (RR) sebagai variabel targetnya. Besarnya intensitas curah hujan dapat dikategorikan menjadi 4 kriteria yang tersaji pada Tabel 2 (Sena et al., 2020) dan Tabel 1 menyajikan sampel data penelitian.

Tabel 1. Sampel Data Penelitian

Tanggal	Tn (°C)	Tx (°C)	Tavg (°C)	RH_avg (%)	SS (jam)	FF_x (m/s)	FF_avg (m/s)	RR (mm)
01-01-2020	26.2	32.6	28.7	79	4.6	5	2	8888
02-01-2020	25.4	34.2	28.9	78	0.3	7	3	1.3
03-01-2020	26	34.2	28.6	81	4	6	3	0
....	....	....	....	....	....	....	....	....
29-12-2024	25.4	31.9	28.1	79	2.5	5	2	12.7
30-12-2024	26	34.2	29.7	69	0.8	5	3	0
31-12-2024	25.6	32.8	28.2	79	5.5	6	3	1

Tabel 2. Kriteria Intensitas Curah Hujan

Kriteria	Intensitas (mm/hari)
Hujan Ringan	5 – 20
Hujan Sedang	20 – 50
Hujan Lebat	50 – 100
Hujan Sangat Lebat	> 100

### Pengisian Data Kosong

Pengisian data kosong dilakukan untuk menambahkan data yang tidak ada untuk memperbaiki kualitas data dan lanjut ke tahap selanjutnya (Alghifari & Juardi, 2021). Terdapat banyak cara untuk melakukan pengisian data kosong, mulai dari menggunakan nilai *mean*, *median*, atau modus dalam data. Cara lainnya seperti interpolasi linier, dan dapat menggunakan nilai sebelumnya atau setelahnya juga dapat dilakukan.

### Normalisasi

Pada suatu *dataset* terdapat kemungkinan data dinilai pada rentang yang berbeda (Khaira et al., 2025). Oleh karena itu, perlu dilakukan normalisasi untuk menghindari kemungkinan terburuk terhadap hasil analisis karena pengaruh rentang data yang tinggi (Kusnanti et al., 2022). Normalisasi merupakan suatu proses yang merubah skala nilai atribut sehingga data memiliki bobot yang sama dalam rentang yang lebih kecil (Pramesti et al., 2022). *Min-max scaler* merupakan salah satu metode normalisasi data yang merubah nilai data menjadi skala 0 sampai 1 dengan melakukan proses transformasi linier (Suryanegara et al., 2021). Normalisasi menggunakan metode *min-max scaler* dapat dihitung menggunakan rumus pada Persamaan (1) (Harlianto et al., 2020).

$$\tilde{x} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Hasil dari normalisasi *min-max* merupakan  $\tilde{x}$ ,  $x_i$  adalah data aktual,  $x_{min}$  merupakan nilai terkecil dari data, dan  $x_{max}$  merupakan nilai terbesar.

### Principal Component Analysis (PCA)

*Principal Component Analysis* (PCA) merupakan metode yang sering diaplikasikan untuk mengurangi dimensi data yang tinggi (Budiarto & Kuntjoro, 2023). PCA dapat memampatkan banyaknya variabel menjadi beberapa variabel komprehensif melalui proses reduksi dimensi (Yang et

al., 2021). PCA juga dapat digunakan untuk mengetahui informasi terkait variabel mana yang memiliki pengaruh terbesar atau bahkan memiliki pengaruh yang sangat kecil (Azmi et al., 2022). Hal ini tentu sangat membantu dalam peningkatan kinerja suatu model.

Langkah pertama dalam PCA adalah menghitung matriks kovarian dengan menggunakan Persamaan (2). Selanjutnya Persamaan (3) berguna untuk menghitung nilai eigen dan Persamaan (4) berguna untuk menghitung nilai vektor eigennya.

$$Cov(x_a, x_b) = \frac{\sum_{i=1}^n (x_{ai} - \bar{x}_a)(x_{bi} - \bar{x}_b)}{n-1} \quad (2)$$

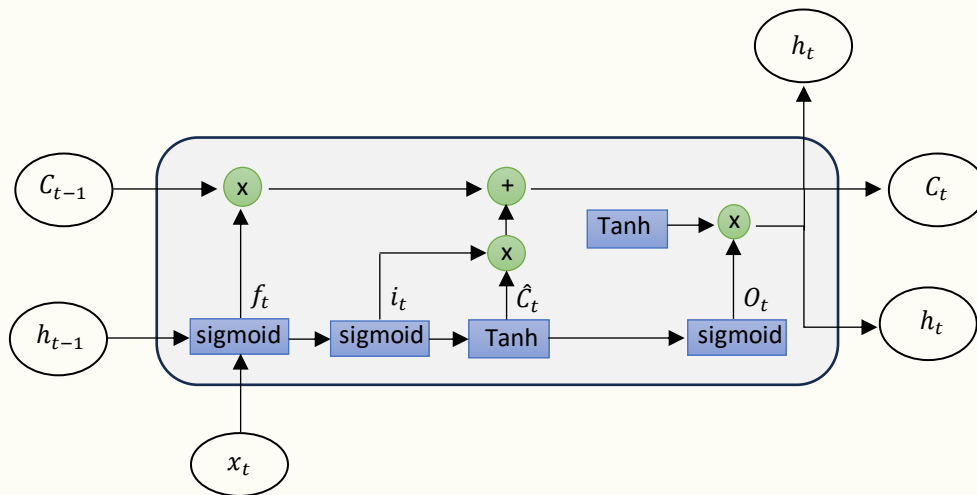
$$\det(A - \lambda I) = 0 \quad (3)$$

$$(A - \lambda I)x = 0 \quad (4)$$

Dimana  $x_a$  dan  $x_b$  merupakan dua data yang berbeda.  $\bar{x}_a$  dan  $\bar{x}_b$  adalah nilai rata-rata  $x_a$  dan  $x_b$ .  $n$  adalah jumlah data.  $A$  adalah sebuah matriks kovarian,  $\lambda$  adalah nilai eigen dari matriks kovarian, dan  $I$  merupakan suatu matriks identitas.

### Long Short-Term Memory

Long Short-Term Memory (LSTM) adalah metode percabangan dari *Recurrent Neural Network* (RNN) yang cocok untuk mengolah data yang berurutan (Rosyd et al., 2024). Metode ini mampu mengatasi masalah gradien meluruh (*vanishing gradient*) dan *long-term dependency* yang dialami oleh RNN sebagai pendahulunya (Wiranda & Sadikin, 2019). Terdapat beberapa lapisan di dalam LSTM yang berfungsi sebagai penyaring informasi, jika informasi relevan maka akan lanjut ke lapisan berikutnya atau sebaliknya jika informasi tidak relevan maka akan dibuang di setiap lapisannya (Hasiholan et al., 2022). Umumnya LSTM memiliki arsitektur yang terdapat pada Gambar 2 (Yang et al., 2021).



Gambar 2. Arsitektur LSTM

LSTM memiliki beberapa *gate* atau gerbang yang digunakan untuk menyaring informasi, seperti *input gate*, *forget gate*, dan *output gate*. *Forget gate* menggunakan *output* di waktu ke  $t - 1$  dan parameter *input* di waktu ke  $t$ . *Forget gate* ( $f_t$ ) menggabungkan kedua parameter tersebut dan mengaplikasikannya pada fungsi aktivasi *sigmoid* ( $\sigma$ ) menggunakan Persamaan (5) yang akan menghasilkan *output* angka 1 atau 0. Informasi yang tidak dihiraukan akan bernilai 0 dan bernilai 1 berarti *state* tidak berubah dan informasinya akan disimpan di *memory cell*. Pada *Input gate* terdapat 2 *gate*, yakni penggunaan fungsi aktivasi *sigmoid* untuk pembaruan informasi pada *memory cell* dan penyimpanan nilai baru dalam *memory cell* menggunakan fungsi dari aktivasi *tanh*. Proses tersebut dapat dilakukan menggunakan Persamaan (6) dan (7).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (5)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (6)$$



$$\hat{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (7)$$

Persamaan (8) digunakan untuk mencari *cell gate* ( $C_t$ ) yang merupakan pembaruan dari *memory cell* yang lama ( $C_{t-1}$ ). Pada *output gate* menggunakan fungsi dari aktivasi *sigmoid* untuk menetapkan nilai yang dihasilkan dan menggunakan fungsi dari aktivasi *tanh* untuk menyimpan suatu nilai tersebut. Proses tersebut dilakukan menggunakan Persamaan (9) dan (10).

$$C_t = f_t * C_{t-1} + i_t * \hat{C}_t \quad (8)$$

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (9)$$

$$h_t = O_t \tanh(C_t) \quad (10)$$

Nilai bobot dari *forget gate*, *input gate*, *cell gate* dan *output gate* dinyatakan dalam  $W_f$ ,  $W_i$ ,  $W_c$ , dan  $W_o$ . Nilai bias dari *forget gate*, *input gate*, *cell gate* dan *output gate* dinyatakan dalam  $b_f$ ,  $b_i$ ,  $b_c$  dan  $b_o$ . Fungsi *sigmoid* dinyatakan dalam  $\sigma$ . Nilai dari *output gate* dinyatakan dalam  $O_t$ . *Output* pada waktu yang ditentukan ( $t$ ) dinyatakan dalam  $h_t$ .

### Denormalisasi

Denormalisasi merupakan proses yang dilakukan setelah dilakukannya pengujian data dengan tujuan untuk mengembalikan data ke rentang aslinya seperti sebelum dilakukannya normalisasi (Mukhtar et al., 2022). Perhitungan denormalisasi dilakukan menggunakan Persamaan (11) (Arfianti et al., 2021).

$$x_i = y(x_{max} - x_{min}) + x_{min} \quad (11)$$

Dimana  $x_i$  adalah data denormalisasi atau data aktual,  $y$  adalah hasil *output*,  $x_{max}$  adalah data asli terbesar, dan  $x_{min}$  adalah data asli terkecil.

### Evaluasi Model

Kinerja suatu model dalam proses prediksi dapat dievaluasi dengan menggunakan metrik yang relevan (Ihzanah et al., 2023). *Mean Arctangent Absolute Percentage Error* (MAAPE) merupakan salah satu metrik yang memiliki kesamaan dengan *Mean Absolute Percentage Error* (MAPE), evaluasi metrik dilakukan dengan mengevaluasi persentase rata-rata dari nilai aktual dan prediksi. Namun, MAAPE menggunakan fungsi *arctangent* sebagai pembatasan terhadap efek *outlier* (Haq et al., 2021). Perhitungan nilai MAAPE dapat menggunakan Persamaan (12).

$$MAAPE = \frac{1}{n} \sum_{i=1}^n \left| \arctan \left( \frac{\tilde{x}_i - x_i}{x_i} \right) \right| \quad (12)$$

Dimana  $n$  merupakan jumlah data,  $x_i$  merupakan data aktual, dan  $\tilde{x}_i$  merupakan data prediksi.

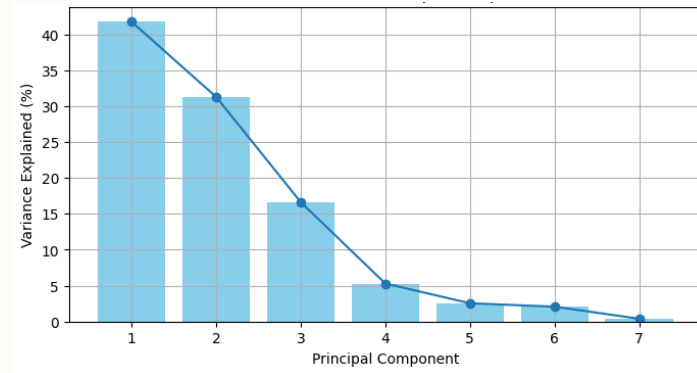
## HASIL DAN PEMBAHASAN

### Preprocessing

Data yang hilang masih terdapat dalam data yang digunakan penelitian ini. Pengisian data kosong tersebut memanfaatkan cara interpolasi dan dimisalkan data yang bernilai 8888 yang berarti data tidak terukur pada variabel curah hujan menjadi 0.1 mm (Musfiroh et al., 2023). Selanjutnya dilakukan proses normalisasi pada variabel fitur dan target menggunakan Persamaan (1). Normalisasi dilakukan untuk menghindari pengaruh rentang data yang terlalu tinggi dan merubah nilai data dalam rentang 0 sampai 1.

### Penerapan Metode PCA

Sebanyak 7 variabel selain variabel curah hujan akan diproses menggunakan *Principal Component analysis* untuk dilakukan reduksi fitur. Hasil proses ini berupa *Principal Component* (PC) yang dilakukan menggunakan Persamaan (2) sampai Persamaan (4).



Gambar 3. Varians Explained setiap PC

Besarnya nilai variance explained juga menjelaskan seberapa besar pengaruh dari setiap PC. Pada Gambar 5, menjelaskan PC1 memiliki nilai variance explained sebesar 41.77% dan PC2 memiliki nilai variance explained sebesar 31.29%. Terdapat juga PC3 sampai PC7 yang memiliki nilai variance explained yang kurang dari 20%. PC1 dan PC2 dapat menjelaskan sebesar 73.06%, sehingga PC1 dan PC2 akan digunakan dalam proses selanjutnya, serta tidak dilakukan analisis lebih lanjut terhadap PC3 sampai PC7.

Tabel 3. Kontribusi Variabel terhadap *Principal Component*

	PC1	PC2
TN	0.328	-0.411
TX	0.454	-0.158
TAVG	0.588	-0.255
RH_AVG	-0.521	-0.263
SS	0.002	-0.003
FF_X	0.062	0.245
FF_AVG	0.255	0.784

Tabel 3 menjelaskan tentang seberapa besar pengaruh masing-masing variabel pada PC1 dan PC2. Nilai kontribusi yang semakin mendekati 1, maka akan semakin besar kontribusinya. Nilai yang bernilai negatif menggambarkan korelasi bernilai negatif juga pada tiap PCnya, begitu juga sebaliknya. Variabel yang memiliki pengaruh besar pada PC1 adalah TAVG dan RH\_AVG, sedangkan variabel yang paling berpengaruh pada PC2 adalah FF\_AVG. Proses PCA menghasilkan variabel PC1 dan PC2 yang merupakan reduksi fitur, didalam variabel ini terdapat semua fitur yang mempengaruhi curah hujan.

### Pola Time Series

Penelitian prediksi curah hujan memakai 3 input, didapat dari variabel curah hujan dan 2 variabel hasil dari PCA untuk memprediksi curah hujan setelahnya. Tabel 4 menjelaskan variabel target dan inputnya diproses.

Tabel 4. Pola Data Input menggunakan PC1, PC2, dan Parameter Curah Hujan

Dataset PC1	Dataset PC2	Dataset Curah Hujan	Target
$PC1_1, PC1_2, \dots, PC1_6$	$PC2_1, PC2_2, \dots, PC2_6$	$RR_1, RR_2, \dots, RR_6$	$RR_7$
$PC1_2, PC1_3, \dots, PC1_7$	$PC2_2, PC2_3, \dots, PC2_7$	$RR_2, RR_3, \dots, RR_7$	$RR_8$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$PC1_{1824}, PC1_{1825}, \dots, PC1_{1829}$	$PC2_{1824}, PC2_{1825}, \dots, PC1_{1829}$	$RR_{1824}, RR_{1825}, \dots, RR_{1829}$	$RR_{1830}$

### Penerapan Metode LSTM

Penentuan nilai parameter seperti *hidden layer*, *batch size*, dan *learn rate drop period* berpengaruh besar terhadap hasil prediksi dari metode LSTM. Pengujian setiap parameter untuk menetapkan nilai terbaik dilakukan dalam 10 kali percobaan dan mengacu pada nilai MAAPE untuk menentukan nilai parameter terbaik dari setiap percobaan. Selain penentuan nilai parameter, pembagian data juga merupakan faktor lain yang dapat mempengaruhi hasil prediksi. Pembagian data untuk uji coba yang dilakukan menggunakan pola data *input* dengan parameter PC1, PC2, dan curah hujan memakai 80% data training menunjukkan nilai MAAPE terkecil dari pembagian data lainnya. Pembagian data diuji coba terhadap setiap parameter untuk menghasilkan model prediksi yang terbaik. Nilai MAAPE terbaik untuk setiap 10 percobaan ditunjukkan pada Tabel 5.

Pada Tabel 5, telah ditunjukkan bahwasannya selisih besarnya MAAPE rata-rata dari ketiga *hidden layer* hanya sedikit berbeda. Nilai MAAPE yang semakin kecil atau mendekati 0, menunjukkan model semakin baik dalam membaca pola data, begitupun sebaliknya. Nilai rata-rata MAAPE terendah ada di *hidden layer* 150, dengan nilai rata-rata 0.7806. Nilai MAAPE terendah pada *hidden layer* 150 terdapat pada *batch size* 256 dan *learn rate drop period* 50, dengan besarnya MAAPE senilai 0.5662, serta besarnya MAAPE tertinggi terdapat pada *batch size* 32 dan *learn rate drop period* 150, dengan besarnya MAAPE senilai 0.9902. Nilai rata-rata MAAPE tertinggi terdapat pada *hidden layer* 50, dengan nilai rata-rata 0.8995. Nilai MAAPE terendah pada *hidden layer* 50 terdapat pada *batch size* 32 dan *learn rate drop period* 100, dengan besarnya MAAPE senilai 0.7878, serta nilai MAAPE tertinggi terdapat pada *batch size* 64 dan *learn rate drop period* 150, dengan besarnya MAAPE senilai 1.0716. Pada *hidden layer* 100, nilai MAAPE terendah terdapat pada *batch size* 128 dan *learn rate drop period* 100, dengan besarnya MAAPE senilai 0.6976, serta nilai MAAPE tertinggi terdapat pada *batch size* 64 dan *learn rate drop period* 100, dengan besarnya 1.0008. Berdasarkan hasil pengujian, model PCA-LSTM mampu membaca pola curah hujan harian dengan cukup baik dengan nilai MAAPE terendah 0.5662 yang disajikan dalam grafik pada Gambar 4.

Tabel 5. Nilai MAAPE dengan Input PC1, PC2, dan RR

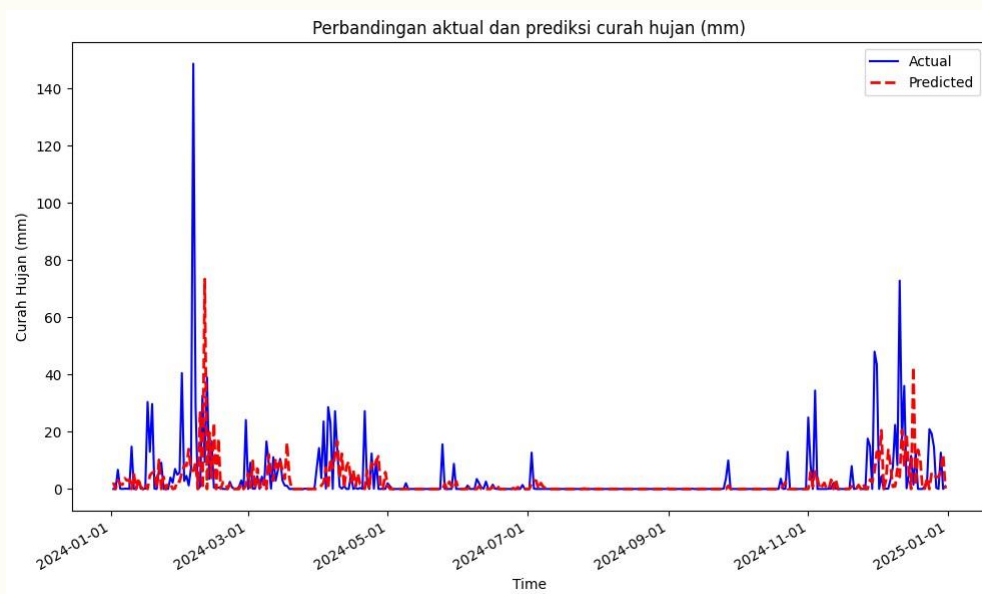
Hidden Layer	Batch Size	Learn Rate Drop Period	MAAPE	Rata-rata MAAPE
50	32	50	0.8659	0.8995
		100	0.7878	
		150	0.9705	
	64	50	0.6879	
		100	0.7887	
		150	1.0716	
	128	50	0.8982	
		100	0.9276	
		150	1.0278	
	256	50	0.8945	
		100	0.9370	
		150	0.9408	
100	32	50	0.7227	0.8245
		100	0.7160	
		150	0.8518	
	64	50	0.8311	
		100	1.0008	
		150	0.7092	
	128	50	0.7864	
		100	0.6976	
		150	0.9264	
	256	50	0.9295	
		100	0.8414	
		150	0.8761	



150	32	50	0.7434	
		100	0.8807	
		150	0.9902	
	64	50	0.6940	
		100	0.7916	
		150	0.7731	
	128	50	0.7065	
		100	0.7343	
		150	0.7203	
	256	50	0.5662	
		100	0.8923	
		150	0.8744	0.7806

### Denormalisasi

Setelah proses prediksi dengan metode *Long Short-Term Memory* (LSTM) dilakukan, data hasil prediksi akan melalui tahap denormalisasi untuk mengembalikan nilai ke skala aslinya agar dapat dibandingkan dengan data aktual. Proses ini penting dilakukan karena model LSTM dapat bekerja lebih optimal pada data yang telah dinormalisasi, namun hasil akhirnya tetap harus dikembalikan ke rentang semula agar grafik prediksi memiliki rentang nilai yang sama dengan data nyata. Denormalisasi membuat interpretasi terhadap hasil prediksi menjadi lebih representatif terhadap kondisi sebenarnya.



Gambar 4. Perbandingan Nilai Prediksi dan Nilai Aktual

Pada Gambar 4 ditunjukkan hasil prediksi curah hujan mulai tanggal 2 Januari 2024 sampai 31 Desember 2024. Grafik tersebut diperoleh dengan menggunakan LSTM dengan menggunakan parameter *hidden layer* 150, *batch size* 256, dan *learn rate drop period* sebesar 50. Grafik tersebut memperlihatkan bahwasannya model PCA-LSTM mampu mengikuti pola umum curah hujan harian, khususnya untuk periode dengan intensitas rendah hingga sedang. Namun model masih cenderung menghasilkan nilai prediksi bernilai lebih rendah dari data aktual. Kondisi ini memperlihatkan bahwa sistem prediksi masih terdapat keterbatasan untuk menangkap pola kejadian ekstrem yang jarang terjadi pada data historis. Hasil prediksi menyatakan puncak curah hujan di wilayah Tanjung Perak Surabaya terjadi pada bulan Februari 2024 dan Desember 2024, hal tersebut sama dengan data aktual dengan nilai prediksi lebih rendah. Curah hujan yang lebat terjadi pada 11 Februari 2024 sebesar 73.3336 mm dan curah hujan sedang terjadi pada 17 Desember 2024 sebesar 42.6660 mm. Perbandingan kuantitatif dengan penelitian terdahulu terdapat pada Tabel 6.

Tabel 6. Gap Penelitian

Peneliti & Tahun	Data	Metode	Jumlah Data	Hasil (MAAPE)
(Haq et al., 2021)	Curah hujan harian Sidoarjo berbasis El-Nino dan IOD.	LSTM	1.680	0.5810
Penelitian ini (2025)	Curah hujan harian Tanjung Perak, Surabaya.	PCA-LSTM	1.830	0.5662

Tabel 6 menunjukkan perbandingan penelitian ini dengan penelitian terdahulu dengan evaluasi model yang sama. Penelitian menggunakan metode LSTM dengan 1.680 data curah hujan harian Sidoarjo berbasis El-Nino dan IOD mendapatkan hasil evaluasi MAAPE 0.5810. Penelitian ini dengan menambahkan metode PCA sebagai reduksi fitur mendapatkan peningkatan hasil akurasi dari penelitian terdahulu. Tabel 6 menunjukkan bahwa penerapan PCA sebagai reduksi fitur sebelum pemodelan LSTM dapat meningkatkan kinerja prediksi.

## KESIMPULAN

Prediksi curah hujan di Tanjung Perak, Surabaya menggunakan algoritma LSTM. Prediksi curah hujan menggunakan data observasi dengan parameter variabel temperatur minimum dan maksimum, temperatur rata-rata, kelembaban rata-rata, lamanya matahari bersinar, kecepatan angin maksimum, dan kecepatan angin rata-rata mulai tanggal 1 Januari 2020 sampai dengan 1 Januari 2024 untuk memprediksi curah hujan 2 Januari 2024 sampai 31 Desember 2024. Variabel tersebut diproses menggunakan PCA untuk melakukan reduksi fitur dan didapatkan bahwa PC1 dan PC2 sudah merepresentasikan 73.06% dari data. PC1 menyatakan variabel rata-rata temperatur dan rata-rata kelembaban merupakan variabel yang paling berpengaruh dan dalam PC2 menyatakan variabel rata-rata kecepatan angin merupakan variabel yang paling berpengaruh. Dengan menggunakan uji coba variasi nilai *hidden layer*, *batch size*, dan *learn rate drop period* didapatkan nilai MAAPE terendah senilai 0.5662 dengan *hidden layer* 150, *batch size* 256, dan *learn rate drop period* 50 yang menunjukkan bahwasannya model yang dihasilkan sudah cukup baik dalam melakukan prediksi pola curah hujan. Meskipun hasil yang diperoleh cukup baik, penelitian ini masih memiliki keterbatasan terutama untuk melakukan prediksi pada pola curah hujan yang ekstrem. Riset ini juga hanya menggunakan data dari satu lokasi sehingga masih perlu dikaji lebih lanjut untuk wilayah lainnya. Penelitian menggunakan metode lain dengan data yang sama juga perlu dikaji lebih lanjut sebagai pembandingan.

## DAFTAR PUSTAKA

- Akbar, R., Santoso, R., & Warsito, B. (2023). Prediksi Tingkat Temperatur Kota Semarang Menggunakan Metode Long Short-Term Memory (LSTM). *Jurnal Gaussian*, 11(4), 572–579. <https://doi.org/10.14710/j.gauss.11.4.572-579>
- Alghifari, F., & Juardi, D. (2021). Penerapan Data Mining Pada Penjualan Makanan dan Minuman Menggunakan Metode Algoritma Naïve Bayes. *Jurnal Ilmiah Informatika*, 9(02), 75–81. <https://doi.org/10.33884/jif.v9i02.3755>
- Amatullah, L., Widiastiwi, Y., & Chamidah, N. (2022). Penerapan Klasifikasi Random Forest Terhadap Data Gangguan Spektrum Autisme (ASD) pada Anak-Anak Menggunakan Seleksi Fitur Principal Component Analysis. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, 3(2), 356–364.
- Arfianti, U. I., Novitasari, D. C. R., Widodo, N., Hafiyusholeh, Moh., & Utami, W. D. (2021). Sunspot Number Prediction Using Gated Recurrent Unit (GRU) Algorithm. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(2), 141-152. <https://doi.org/10.22146/ijccs.63676>

- Ashari, M. L., & Sadikin, M. (2020). Prediksi Data Transaksi Penjualan Time Series Menggunakan Regresi Lstm. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 9(1), 1-10. <https://doi.org/10.23887/janapati.v9i1.19140>
- Azmi, B. N., Hermawan, A., & Avianto, D. (2022). Analisis Pengaruh PCA pada Klasifikasi Kualitas Air Menggunakan Algoritma K-Nearest Neighbor dan Logistic Regression. *Jurnal Sistem Dan Teknologi Informasi*, 7(2), 94–103.
- Budiarto, R., & Kuntjoro, Y. D. (2023). Analisis Perilaku Entitas untuk Pendeteksian Serangan Internal Menggunakan Kombinasi Model Prediksi Memori dan Metode PCA. 10(6), 1223–1232. <https://doi.org/10.25126/jtiik.2023107123>
- Fang, Q., Zhong, Y., Xie, C., Zhang, H., & Li, S. (2020). Research on PCA-LSTM-based Short-term Load Forecasting Method. *IOP Conference Series: Earth and Environmental Science*, 495(1). <https://doi.org/10.1088/1755-1315/495/1/012015>
- Firdaus, A. J. A., Pramono, D., & Purnomo, W. (2020). Pengembangan Sistem Informasi UPT Kalibrasi Dinas Kesehatan Kabupaten Malang Berbasis WEB. *Jurnal Sistem Informasi, Teknologi Informasi, Dan Edukasi Sistem Informasi*, 1(1), 23–34. <https://doi.org/10.25126/justsi.v1i1.3>
- Haq, D. Z., Rini Novitasari, D. C., Hamid, A., Ulinnuha, N., Arnita, Farida, Y., Nugraheni, R. D., Nariswari, R., Ilham, Rohayani, H., Pramulya, R., & Widjayanto, A. (2021). Long Short-Term Memory Algorithm for Rainfall Prediction Based on El-Nino and IOD Data. *Procedia Computer Science*, 179(2019), 829–837. <https://doi.org/10.1016/j.procs.2021.01.071>
- Harlianto, D., Mardiyati, S., Lestari, D., Zili, A. H., & Devila, S. (2020). Indonesia Tuberculosis Morbidity Rate Forecasting Using Recurrent Neural Network. *AIP Conference Proceedings*, 2242(June 2020), 030006. <https://doi.org/10.1063/5.0010445>
- Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. *Journal of Applied Informatics and Computing*, 5(2), 103–108. <https://doi.org/10.30871/jaic.v5i2.3200>
- Hasiholan, A., Cholissodin, I., & Yudistira, N. (2022). Analisis Sentimen Tweet Covid-19 Varian Omicron pada Platform Media Sosial Twitter menggunakan Metode LSTM berbasis Multi Fungsi Aktivasi dan GLOVE. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 6(10), 4653–4661.
- Herdhyanti, A., Muflikhah, L., & Cholissodin, I. (2022). Prediksi Curah Hujan dengan Empat Parameter menggunakan Backpropagation (Studi Kasus: Stasiun Meteorologi Ahmad Yani). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 6(12), 5862–5870.
- Hidayatullah, S., & Cherid, A. (2023). Prediksi Temperatur Cuaca di Negara Norwegia Menggunakan Metode LSTM. *Jurnal Sistem Informasi dan Sistem Komputer*, 8(2), 187–198. <https://doi.org/10.51717/simkom.v8i2.192>
- Ihzaniah, L. S., Setiawan, A., & Wijaya, R. W. N. (2023). Perbandingan Kinerja Metode Regresi K-Nearest Neighbor dan Metode Regresi Linear Berganda pada Data Boston Housing. *Jambura Journal of Probability and Statistics*, 4(1), 17–29. <https://doi.org/10.34312/jjps.v4i1.18948>
- Istianto, A. F., Hadiana, A. I., & Umbara, F. R. (2024). Prediksi Curah Hujan Menggunakan Metode Categorical Boosting (CatBoost). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(4), 2930–2937. <https://doi.org/10.36040/jati.v7i4.7304>
- Khaira, U., Putri, M. F., & Yanova, S. (2025). Peramalan Kadar PM 10 Menggunakan Algoritma Long Short-Term Memory (LSTM) Sebagai Acuan Ketersediaan Ruang Terbuka Hijau Di Kota Jambi. *Decode: Jurnal Pendidikan Teknologi Informasi*, 5(1), 289–302. <https://doi.org/http://dx.doi.org/10.51454/decode.v5i1.866>

- Kosasih, A., Hartono, H., & Jatmiko, R. H. (2021). Pengaruh Koreksi Atenuasi Radar Cuaca Terhadap Perhitungan Estimasi Curah Hujan Di Jawa Timur. *Jurnal Teknosains*, 10(2), 111-124.
- Kusnanti, E. A., Novitasari, D. C. R., Setiawan, F., Fanani, A., Hafiyusholeh, M., & Sari, G. I. P. (2022). Predicting Velocity and Direction of Ocean Surface Currents using Elman Recurrent Neural Network Method. *Journal of Information Systems Engineering and Business Intelligence*, 8(1), 21-30. <https://doi.org/10.20473/jisebi.8.1.21-30>
- Kusnanti, E. A., Vantie, L. D. F., & Yuhana, U. L. (2024). Software Defect Prediction Using Pca Based Recurrent Neural Network. *JUTI: Jurnal Ilmiah Teknologi Informasi*, 22(1), 23-31.
- Luthfiarta, A., Febriyanto, A., Lestiawan, H., & Wicaksono, W. (2020). Analisa Prakiraan Cuaca dengan Parameter Suhu, Kelembaban, Tekanan Udara, dan Kecepatan Angin Menggunakan Regresi Linear Berganda. *JOINS (Journal of Information System)*, 5(1), 10-17. <https://doi.org/10.33633/joins.v5i1.2760>
- Merdekawati, G. I., & Ismail. (2019). Prediksi Curah Hujan Di Jakarta Berbasis Algoritma Levenberg Marquardt. *Jurnal Ilmiah Informatika Komputer*, 24(2), 116-128.
- Mukhtar, H., Gunawan, R., Hariyanto, A., Syahril, & Mulyana. W. (2022). Peramalan Kedatangan Wisatawan ke Suatu Negara Menggunakan Metode Support Vector Machine (SVM). *Jurnal CoSciTech (Computer Science and Information Technology)*, 3(3), 274-282. <https://doi.org/10.37859/coscitech.v3i3.4211>
- Musfiroh, M., Novitasari, D. C. R., Intan, P. K., & Wisnawa, G. G. (2023). Penerapan Metode Principal Component Analysis (PCA) dan Long Short-Term Memory (LSTM) dalam Memprediksi Prediksi Curah Hujan Harian. *Building of Informatics, Technology and Science (BITS)*, 5(1), 1-11. <https://doi.org/10.47065/bits.v5i1.3114>
- Peranginangin, I., Zakaria, A., & Kusumastuti, D. I. (2021). Perbandingan spektrum curah hujan harian antara BMKG dan TRMM di DKI Jakarta. *REKAYASA: Jurnal Ilmiah Fakultas Teknik Universitas Lampung*, 24(2), 53-57.
- Pramesti, D. D., Novitasari, D. C. R., Setiawan, F., & Khaulasari, H. (2022). Long-Short Term Memory (LSTM) for Predicting Velocity and Direction Sea Surface Current on Bali Strait. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 16(2), 451-462.
- Ramadhan, F., & Fachrie, M. (2024). Implementasi Algoritma Long Short-Term Memory Pada Sistem Prediksi Hasil Panen Sawit. *Jurnal Informatika Teknologi Dan Sains (Jinteks)*, 6(4), 937-944. <https://doi.org/10.51401/jinteks.v6i4.4876>
- Rizkilloh, M. F., & Widiyanesti, S. (2022). Prediksi Harga Cryptocurrency Menggunakan Algoritma Long Short Term Memory (LSTM). *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 6(1), 25-31. <https://doi.org/10.29207/resti.v6i1.3630>
- Rosyd, A., Purnamasari, A. I., & Ali, I. (2024). Penerapan Metode Long Short Term Memory (Lstm) Dalam Memprediksi Harga Saham Pt Bank Central Asia. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 501-506. <https://doi.org/10.36040/jati.v8i1.8440>
- Salehe, S., Musa, R., & Sar, M. (2024). Kajian Curah Hujan Akibat Pengaruh Temperatur, Kelembaban dan Kecepatan Angin (Studi Kasus Stasiun Klimatologi Bonto Bili Kab. Gowa). *Jurnal Teslink : Teknik Sipil Dan Lingkungan*, 6(2), 458-467. <https://doi.org/10.52005/teslink.v6i2.372>
- Satyo, A., Karno, B., & Ali, J. K. N., (2020). Prediksi Data Time Series Saham Bank BRI Dengan Mesin Belajar LSTM (Long ShortTerm Memory). *Journal of Informatic and Information Security*, 1(1), 2722-4058.

- Sena, I. G. W., Dillak, J. W., Leunupun, P., & Santoso, A. J. (2020). Predicting Rainfall Intensity using Naïve Bayes and Information Gain Methods (Case Study: Sleman Regency). *Journal of Physics: Conference Series*, 1577(1). <https://doi.org/10.1088/1742-6596/1577/1/012011>
- Sunarmi, N., Kumailia, E. N., Nurfaiza, N., Nikmah, A. K., Aisyah, H. N., Sriwahyuni, I., & Lailly, S. N. (2022). Analisis Faktor Unsur Cuaca terhadap Perubahan Iklim Di Kabupaten Pasuruan pada Tahun 2021 dengan Metode Principal Component Analysis. *Newton-Maxwell Journal of Physics*, 3(2), 56–64.
- Suryanegara, G. A. B., Adiwijaya, A., & Purbolaksono, M. D. (2021). Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi Pasien Penderita Diabetes Menggunakan Metode Normalisasi. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(1), 114–122. <https://doi.org/10.29207/resti.v5i1.2880>
- Wiranda, L., & Sadikin, M. (2019). Penerapan Long Short Term Memory pada Data Time Series untuk Memprediksi Penjualan Produk PT. Metiska Farma. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 8(3), 184–196.
- Yang, K., Yuan, J. L., Xiong, T., Wang, B., & Fan, S. D. (2021). A Novel Principal Component Analysis Integrating Long Short-Term Memory Network And Its Application In Productivity Prediction Of Cutter Suction Dredgers. *Applied Sciences (Switzerland)*, 11(17), 8159. <https://doi.org/10.3390/app11178159>