

# Rainfall Forecasting Using SSA-Based Hybrid Models with LSSVR and LSTM for Disaster Mitigation

Zauyik Nana Ruslana\*<sup>1</sup>, Eri Zuliarso<sup>2</sup>

<sup>1,2</sup>Master of Information Technology, Faculty of Information Technology and Industry, Stikubank University, Indonesia

Email: <sup>1</sup>zauyiknana0024@mhs.unisbank.ac.id

Received : Jun 27, 2025; Revised : Jul 13, 2025; Accepted : Jul 14, 2025; Published : Aug 18, 2025

## Abstract

Accurate rainfall forecasting is crucial for addressing the increasing risk of hydrometeorological disasters, particularly in tropical regions such as Semarang City, Indonesia. However, conventional forecasting models often struggle with inaccurate data and observations. This study proposes a novel hybrid combination of SSA-NMF with LSSVR and LSTM, offering high-resolution rainfall forecasting over multiple monitoring stations, to predict daily rainfall. As a preprocessing step, 15 years of daily rainfall data from six observation stations were denoised and decomposed using Singular Spectrum Analysis (SSA) combined with Non-Negative Matrix Factorization (NMF). This approach effectively handled data with many zero values, identified seasonal patterns or high-rainfall locations, and extracted key patterns. The prediction models were trained and validated using parameters optimized through RandomizedSearchCV for LSSVR and Keras Tuner for LSTM. Model performance was evaluated using MSE, RMSE, MAE, and Nash-Sutcliffe Efficiency (NSE). The results showed that the SSA-LSTM model consistently outperformed SSA-LSSVR model, with the highest average NSE value being 0.9 across six monitoring locations in Semarang City. Furthermore, the predicted rainfall values were spatially visualized using Inverse Distance Weighting (IDW) interpolation within a Geographic Information System (GIS) environment, producing informative rainfall distribution maps that support early warning systems and disaster mitigation efforts. In conclusion, the hybrid approach combining SSA-NMF preprocessing with LSTM-based deep learning significantly improves the accuracy and reliability of daily rainfall forecasting. This novel SSA-NMF + LSSVR/LSTM framework delivers high-resolution, reliable rainfall forecasts that directly empower disaster risk reduction systems and readily transfer to similar climatic regions.

**Keywords :** LSSVR, LSTM, NMF, Rainfall Prediction, Spatial Visualization, SSA.

This work is an open access article and licensed under a Creative Commons Attribution-Non Commercial 4.0 International License



## 1. PENDAHULUAN

Negara kepulauan seperti Indonesia yang memiliki iklim tropis, di mana curah hujan tinggi dan pola cuaca yang kompleks. Banyak faktor meteorologis dan iklim, seperti ENSO (*El Nino Southern Oscillation*) dan IOD (*Indian Ocean Dipole*), berkontribusi pada curah hujan. Ketidakpastian dalam curah hujan berpotensi menyebabkan bencana hidrometeorologi terkait air, sehingga prediksi yang tepat sangat krusial untuk perencanaan mitigasi bencana. Penggunaan kecerdasan buatan dalam prediksi curah hujan semakin meningkat. Model-model seperti LSTM (*Long Short Term Memory*) dan LSSVR (*Least Square Support Vector Regression*) terbukti efektif dalam mendeteksi pola yang rumit, meskipun masih menghadapi masalah terkait fluktuasi yang tidak stabil. Metode hybrid yang menggabungkan SSA (*Singular Spectrum Analysis*), LSSVR, dan LSTM dapat meningkatkan ketepatan prediksi. Pemahaman distribusi curah hujan juga sangat bergantung pada visualisasi spasial. Menggunakan teknik GIS (*Geographic Information System*) untuk menyusun peta spasial dapat mendukung proses pengambilan keputusan oleh Stasiun Klimatologi Jawa Tengah serta para pemangku kepentingan lainnya. Tujuan dalam penelitian ini adalah untuk menciptakan model hybrid guna prediksi curah hujan

dan menerapkan visualisasi spasial untuk mitigasi bencana. Diharapkan hasil yang diperoleh dapat meningkatkan ketepatan prakiraan dan memberikan informasi yang lebih jelas. Studi kasus dilakukan di Kota Semarang yang memiliki kondisi topografi dataran rendah, berdekatan dengan laut, serta area dataran tinggi.

Pengolahan, analisa dan prediksi dengan pendekatan SSA-LSSVR dan SSA-LSTM belum pernah dibahas secara mendalam sebelumnya dan masih jarang dalam literatur ilmiah, serta mengetahui bagaimana proses pengolahan dan analisis data curah hujan menggunakan kombinasi metode tersebut. Penelitian ini untuk mengetahui seberapa baik performa metode hybrid SSA-LSSVR dan SSA-LSTM dalam memprediksi curah hujan harian dan bulanan di Kota Semarang, berdasarkan metrik evaluasi kuantitatif seperti RMSE (*Root Mean Square Error*), MSE (*Mean Squared Error*), MAE (*Mean Absolute Error*), dan NSE (*Nash-Sutcliffe Efficiency*). Kemudian untuk mengetahui model hybrid mana, antara SSA-LSSVR atau SSA-LSTM, yang memberikan akurasi terbaik dalam prediksi curah hujan di Kota Semarang berdasarkan hasil evaluasi kinerja model. Selanjutnya dilakukan visualisasi spasial hasil prakiraan curah hujan yang dihasilkan oleh model terbaik, serta bagaimana distribusi curah hujan tersebut menggambarkan kondisi hidroklimatologi di wilayah Kota Semarang. Langkah selanjutnya, mengetahui implikasi penggunaan model hybrid ini terhadap pengembangan sistem peringatan dini bencana hidrometeorologi berbasis prediksi curah hujan di tingkat lokal.

Penelitian ini berupaya mengembangkan model hybrid SSA-LSSVR dan SSA-LSTM dalam prakiraan curah hujan harian di Kota Semarang di masa depan dapat dikembangkan lagi ke wilayah lain di Jawa Tengah dengan perbaikan dan penelusuran metode terbaiknya untuk output prediksi selain harian yaitu dasarian, bulanan dan tahunan. Masih sedikitnya literatur yang membahas tentang metode hybrid SSA-LSSVR apalagi untuk model hybrid SSA-LSTM masih belum ada, kebanyakan hanya melakukan penelitian perbagian model saja yaitu hanya SSA atau hanya tentang LSTM untuk analisa dan prediksi curah hujan. Selanjutnya membandingkan performa kedua metode berdasarkan evaluasi statistik dan memetakan hasil prediksi curah hujan menggunakan teknik interpolasi IDW (*Inverse Distance Weighted*) untuk analisis distribusi spasialnya.

Bencana hidrometeorologi merupakan ancaman yang semakin sering terjadi seiring dengan meningkatnya perubahan iklim dan aktivitas manusia. Bencana ini muncul dari interaksi kompleks antara atmosfer, hidrosfer, dan litosfer, serta berkaitan erat dengan curah hujan ekstrem yang berdampak langsung pada kehidupan sosial, ekonomi, dan lingkungan [1],[2]. Beberapa dekade terakhir, prediksi curah hujan menjadi tantangan besar dalam ilmu cuaca dan iklim karena sifatnya yang kompleks, tidak linier, serta mengandung ketidakpastian tinggi [3]. Oleh sebab itu, dibutuhkan pendekatan pemodelan yang cermat, mulai dari prapemrosesan data, deteksi pola, hingga penggunaan teknologi kecerdasan buatan.

Salah satu teknik prapemrosesan yang kini menjadi perhatian utama adalah *Singular Spectrum Analysis* (SSA), yang terbukti efektif dalam memisahkan tren, musiman, dan noise pada data deret waktu [4],[5]. SSA mampu menangani noise secara efisien tanpa asumsi distribusi data, sehingga cocok untuk data klimatologis yang seringkali fluktuatif dan tidak stabil. Untuk meningkatkan akurasi prediksi, SSA sering digabungkan dengan metode lain dalam skema *hybrid modeling*, seperti yang dilakukan oleh Olushina Olowale Awe [6] melalui kombinasi SSA dan *Non-negative Matrix Factorization* (NMF) guna merekonstruksi data curah hujan yang non-negatif secara lebih akurat.

Dalam perkembangan terkini, integrasi SSA dengan teknik pembelajaran mesin seperti *Genetic Algorithm* (GA), *Support Vector Machine* (SVM), dan *Long Short-Term Memory* (LSTM) menunjukkan hasil sangat menjanjikan. Misalnya, penelitian Anh Duy Nguyen [7] memanfaatkan kombinasi SSA dan GA untuk prakiraan debit sungai, dengan peningkatan performa prediksi yang signifikan hingga 40% berdasarkan nilai NSE. Pendekatan hybrid SSA-SVM juga berhasil diterapkan untuk prediksi harga listrik jangka pendek dengan optimasi fitur melalui algoritma *Cuckoo* [8]. Demikian juga model SSA-

SVR mengungguli metode konvensional seperti STL, ARMA, dan MLR dalam mendeteksi tren kegagalan data deret waktu [9].

Lebih lanjut, menggabungkan SSA dengan *Least Squares Support Vector Machine* (LSSVM) dan *Deep Belief Network* (DBN) untuk prakiraan tenaga angin jangka pendek, menunjukkan error yang sangat rendah dengan RMSE 2.13% [10]. Kombinasi SSA, *Jordan Neural Network*, dan *Echo State Network* mampu memperkirakan beban dan harga listrik dengan MAPE di bawah 2%, jauh lebih baik dibandingkan SSA tunggal [11].

Dalam konteks prediksi curah hujan, model hybrid SSA-LSSVR-RF terbukti memberikan kinerja yang solid [3], kemudian keandalan *Support Vector Regression* (SVR) dalam berbagai domain aplikasi karena kemampuannya menangani generalisasi data [12]. Penelitian yang dilakukan oleh Neeraj [13] semakin memperkuat argumen tersebut dengan menunjukkan bahwa model SSA-LSTM memberikan performa prediksi beban listrik yang lebih baik dibandingkan metode populer lain seperti ARMAX, RF, ANN, dan EMD-SVR.

Penelitian Jian Zhao [14] yaitu mengintegrasikan SSA dan LSTM untuk memodelkan tren kenaikan permukaan laut di Laut Kuning. Hasilnya menunjukkan peningkatan akurasi prediksi yang signifikan dengan RMSE minimum mencapai 35.04 mm dibandingkan dengan model LSTM tunggal.

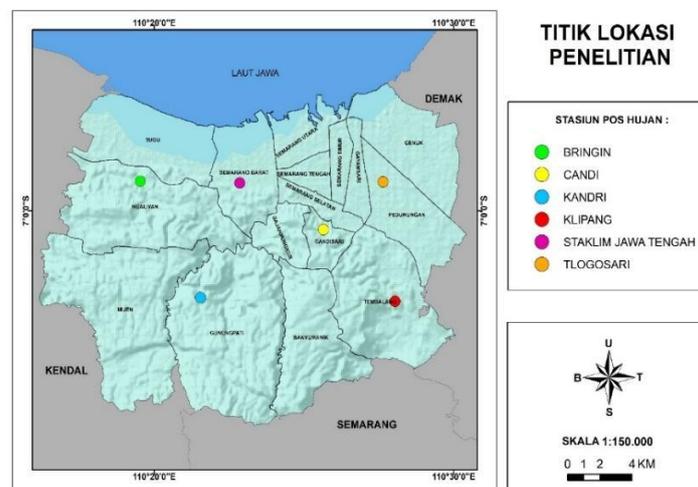
Berbeda dengan penelitian sebelumnya [15],[16] yang menerapkan LSTM dan penelitian [17] yang menerapkan LSSVR secara langsung pada data series curah hujan. Penelitian ini memperkenalkan SSA-NMF sebagai strategi dekomposisi sebelum pemodelan, yang secara signifikan meningkatkan ekstraksi pola dan penyaringan noise. Dengan demikian, penggunaan metode hybrid yang memanfaatkan kekuatan SSA dalam prapemrosesan data dan algoritma pembelajaran mesin dalam prediksi tidak hanya meningkatkan akurasi, tetapi juga memberikan solusi adaptif dalam menghadapi bencana hidrometeorologi di masa mendatang.

Penelitian ini menawarkan kontribusi baru dengan mengintegrasikan SSA dengan LSSVR dan LSTM, menggunakan NMF pada tahap praproses, yang jarang diterapkan dalam prakiraan curah hujan harian.

## 2. METODE

Ringkasan tahapan metodologi penelitian pada tabel 1, memberikan gambaran diskriptif tahapan dari awal sampai akhir merangkum seluruh alur kerja penelitian secara sistematis. Tabel ini menunjukkan setiap langkah, deskripsi singkatnya, dan parameter kunci yang digunakan atau di-tuning.

### 2.1. Data dan Lokasi Penelitian



Gambar 1. Titik lokasi penelitian

Tabel 1. Tahapan metodologi

| Tahapan               | Deskripsi  | Parameter / Metode yang Digunakan   |
|-----------------------|--|---|
| Persiapan Data        | Bentuk data curah hujan harian dikumpulkan dari 6 pos pengamatan selama 15 tahun di Kota Semarang. Melakukan penanganan data hilang (imputasi data) dan memastikan konsistensi data.   | Sumber: Data harian (2010-2024) dari 6 pos hujan. Metode: Interpolasi linear.   |
| Dekomposisi Sinyal    | Menguraikan sinyal curah hujan asli menggunakan Singular Spectrum Analysis (SSA) untuk memisahkan komponen tren, musiman, dan noise. Dekomposisi data untuk mengekstrak tren dan pola utama dengan Non-Negative Matrix Factorization (NMF) karena bekerja baik pada data dengan banyak nilai nol dan dapat mengidentifikasi pola musiman atau lokasi dengan curah hujan tinggi, selain itu hasil dekomposisi NMF selalu non-negatif, sehingga lebih interpretable untuk data curah hujan | Metode: SSA - NMF Parameter di-tuning: Panjang Jendela (Window size), komponen numerik, matrik koefisien dan matrik basis. Nilai panjang jendela dipilih berdasarkan karakteristik data untuk menangkap periodisitas utama yaitu siklus triwulan atau 90 hari.  |
| Rekonstruksi Komponen | Mengelompokkan komponen hasil dekomposisi SSA-NMF menjadi sinyal yang lebih bermakna (misalnya, gabungan Tren + Musiman) yang akan digunakan sebagai input untuk model prakiraan.  | Output: Komponen sinyal yang direkonstruksi (yaitu: Sinyal Tren, Sinyal Musiman).   |
| Optimisasi & Training | Melatih model peramalan menggunakan komponen hasil SSA-NMF sebagai input. Hyperparameter dari setiap model di-tuning untuk mendapatkan performa terbaik.   | <b>Model 1: Hybrid SSA-LSSVR</b><br>Tuning: RandomizedSearchCV<br>Hyperparameter: Parameter ( $\epsilon$ , $C$ , dan $\gamma$ ) & Kernel (Linear, RBF)<br><b>Model 2: Hybrid SSA-LSTM</b><br>Tuning: pencarian sistematis dengan Keras Tuner -<br>Hyperparameter: Jumlah unit, learning rate, jumlah lapisan, dropout rate, batch size. |
| Evaluasi Model        | Mengevaluasi performa model peramalan terbaik pada data uji menggunakan metrik evaluasi standar untuk tugas regresi.   | Metrik: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Square Error (MAPE) dan Nash Sutcliffe Efficiency (NSE)  |
| Prakiraan             | Menggunakan model yang berperforma terbaik dengan hyperparameter optimal yang telah dilatih untuk melakukan prediksi curah hujan harian.   | Output: Nilai prakiraan curah hujan harian untuk 10 hari mendatang.   |
| Visualisasi Spasial   | Memetakan hasil prediksi dari model berperforma terbaik pada GIS   | Output: pemetaan dengan GIS dengan metode IDW agar menghasilkan peta spasial yang informatif  |

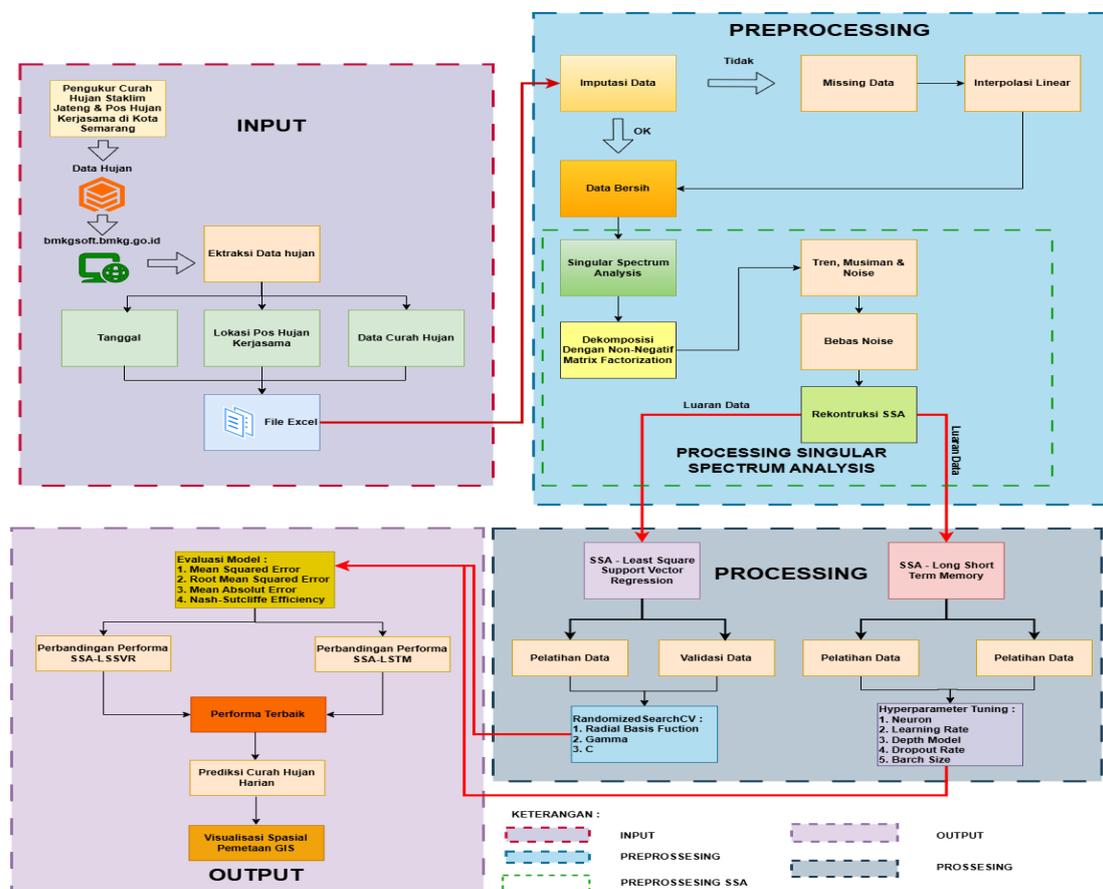
Data yang diambil berasal stasiun pos curah hujan yang tersebar di 6 (enam) titik lokasi di Kota Semarang yaitu Stasiun Klimatologi Jawa Tengah, Kandri, Candi, Bringin, Klipang dan Tlogosari. Data curah hujan adalah data harian yang dikumpulkan selama 15 tahun terakhir, yakni dari tahun 2010 hingga 2024 dari Stasiun Klimatologi Jawa Tengah. Data curah hujan tersebut berasal dari sistem aplikasi berbasis web milik BMKG untuk mengelola dan memantau data meteorologi, klimatologi, dan geofisika secara online dan terpusat (*BMKGsoft*).

Tabel 2. Informasi alat pengukur curah hujan di wilayah penelitian

| Nomor Pos Hujan | Nama Pos Hujan      | Bujur  | Lintang | Elevasi (m) |
|-----------------|---------------------|--------|---------|-------------|
| 33741604a       | Bringin             | 110.33 | -6.99   | 53          |
| 33740702a       | Candi               | 110.43 | -7.01   | 16          |
| 33741205a       | Kandri              | 110.36 | -7.05   | 168         |
| 33741005a       | Klipang             | 110.47 | -7.05   | 91          |
| 96835           | Staklim Jawa Tengah | 110.38 | -6.99   | 5           |
| 33740605a       | Tlogosari           | 110.46 | -6.98   | 5           |

## 2.2. Tahapan Penelitian

Penelitian ini melalui empat tahapan proses yaitu: input, preprocessing, processing dan output. Proses tersebut terlihat pada gambar 1. Penjelasan per tahap dalam penelitian ini terdapat pada sub bab berikutnya.



Gambar 2. Diagram alur tahapan penelitian prediksi curah hujan harian dengan pendekatan hybrid SSA-LSSVR dan SSA-LSTM di Kota Semarang

### 2.3. Input

Ekstraksi data curah hujan harian yang tercatat dalam satu hari dengan satuan milimeter dengan menggunakan alat pengukur curah hujan manual atau *ombrometer* dari Stasiun Klimatologi Jawa Tengah, yang bertugas mengumpulkan data iklim spesifik untuk wilayah Jawa Tengah dan beberapa pos hujan di Kota Semarang kemudian di entry ke *BMKGsoft*. Ekstraksi data dalam bentuk format tabel runtut waktu (*time series*) pada *microsoft excel*. Variabel yang diambil mencakup tanggal, tempat stasiun observasi hujan, dan total curah hujan harian dalam satuan milimeter. Pengolahan data excel tersebut sebagai input pertama yang selanjutnya dilakukan *preprocessing* dan menghasilkan input dengan menggunakan pemrograman *python*.

### 2.4. Preprocessing

#### 2.4.1. Imputasi Data

Tahap *preprocessing* dimulai dengan penanganan data hilang yang kerap muncul pada pengukuran curah hujan akibat gangguan alat atau faktor teknis lain. Karena urutan data yang utuh sangat penting untuk analisis deret waktu, digunakan interpolasi linier sebagai metode *univariate single imputation* untuk mengisi kekosongan secara sistematis dan menjaga kontinuitas pola musiman [18], [19]. Setelah itu, data dinormalisasi dengan *min-max normalization* agar skala variabel seimbang tanpa mengubah hubungan internal data, yang pada akhirnya meningkatkan stabilitas dan kecepatan konvergensi model [20], [21].

#### 2.4.2. Dekomposisi SSA-NMF

Data yang telah bersih dan terstruktur diteruskan ke *Singular Spectrum Analysis* (SSA). Pemakaian SSA dijadikan sebagai teknik pra-pemrosesan untuk membongkar sinyal data curah hujan yang rumit dan banyak noise menjadi bagian yang mudah dipahami dan dapat dianalisis guna memisahkan deret waktu menjadi elemen tren dan musiman. Hasil rekonstruksi SSA yang bebas dari fluktuasi acak kemudian dijadikan input dalam model hybrid SSA-LSTM atau SSA-LSSVR. Pendekatan model hybrid dirancang untuk menghasilkan prakiraan curah hujan yang lebih akurat dan andal, sekaligus aplikatif untuk mitigasi bencana hidrometeorologi [22]. Berikut ini rumus dan langkah-langkahnya dengan teknik analisis deret waktu menggunakan SSA secara umum [23]:

1. *Embedding* :

Mengubah deret waktu  $X = (x_1, x_2, \dots, x_n)$  menjadi matrik H (*Hankel Matrix*) dengan dimensi  $L \times K$  Dimana L adalah *window length* dan N adalah Panjang deret waktu asli dan  $K = N - L + 1$

$$H = [x_1 \ x_2 \ \dots \ x_K \ x_2 \ x_3 \ \dots \ x_{K-1} \ \vdots \ \vdots \ \ddots \ \vdots \ x_L \ x_{L+1} \ \dots \ x_N] \quad (1)$$

2. *Singular Value Decomposition* (SVD) :

Melakukan dekomposisi SVD pada matrik H

$$H = USV^T \quad (2)$$

Keterangan formula :

- U adalah matriks orthogonal  $L \times L$  (vektor singular bagian kiri)
  - S adalah matriks diagonal  $L \times K$  (nilai singular)
  - V adalah matriks orthogonal  $K \times K$  (vektor singular bagian kanan)
3. *Grouping* merupakan proses pengelompokan komponen SVD yang berkaitan dengan tren, musiman, atau noise.
4. *Rekontruksi* deret waktu asli dari komponen yang dipilih menggunakan metode *diagonal averaging* mengubah *matriks Hankel* (juga dikenal sebagai matriks hasil dekomposisi SVD)

kembali menjadi deret waktu dengan menggunakan nilai rata-rata sepanjang diagonal antidiagonal atau diagonal sekunder. *Diagonal averaging* dilakukan dengan cara berikut :

- a. Identifikasi diagonal dimana setiap *diagonal antidiagonal* pada matriks H direpresentasikan oleh indeks  $i + j = konstanta$  Dimana  $i$  adalah indeks baris dan  $j$  adalah indeks kolom.
- b. Menghitung rata-rata diagonal dimana untuk setiap diagonal , dengan menghitung rata-rata sepanjang diagonal tersebut. Nilai rata-rata diagonal akan menjadi nilai deret waktu pada posisi tertentu, formula umum *diagonal averaging* adalah sebagai berikut :

$$x_k = \frac{1}{n_k} \sum_{(i,j) \in D_k} h_{i,j} \tag{3}$$

Keterangan formula :

- $x_k$  nilai deret waktu pada posisi  $k$
  - $h_{i,j}$  elemen matriks H baris  $i$  dan kolom  $j$
  - $D_k$  himpunan indeks  $(i, j)$  yang memenuhi  $i + j - 1 = k$
  - $n_k$  jumlah elemen pada diagonal  $D_k$
- c. Rekonstruksi deret waktu yaitu setelah menghitung rata-rata semua diagonal, maka deret waktu  $X = (x_1, x_2, \dots, x_n)$  akan direkonstruksi.

Namun dalam penelitian ini pada salah satu tahapan proses dekomposisi menggunakan NMF sebagai pendekatan alternatif untuk SVD yang mengasumsikan data dan komponennya bersifat non-negatif [6]. Dekomposisi data untuk mengekstrak tren dan pola utama dengan *Non-Negative Matrix Factorization* (NMF) karena NMF bekerja baik pada data dengan banyak nilai nol dan dapat mengidentifikasi pola musiman atau lokasi dengan curah hujan tinggi, selain itu hasil dekomposisi NMF (matriks basis dan koefisien) selalu non-negatif, sehingga lebih interpretable untuk data curah hujan. Formula dari NMF adalah sebagai berikut :

- a. Matriks non-negatif V (berukuran  $m \times n$ ), NMF mencari dua matriks non-negatif W (berukuran  $m \times k$ ) dan H (berukuran  $k \times n$ ) sehingga :

$$V \approx WH \tag{4}$$

Keterangan formula :

- W adalah matriks basis (*basis vectors*)
  - H adalah matriks koefisien (*coefficient matrix*)
  - $k$  adalah jumlah komponen (dipilih oleh pengguna)
- b. Optimasi NMF dengan tujuan meminimalkan fungsi objektif, dengan menggunakan *Frobenius norm*, dengan formula :

$$W, H \min \|V - WH\|_F^2 \tag{5}$$

Dengan batasan  $W \geq 0$  dan  $H \geq 0$

- c. Algoritma NMF yaitu *Multiplicative Update Rule* dengan formula :

$$W_{i,j} \leftarrow W_{i,j} \frac{(vH^T)_{ij}}{(WHH^T)_{ij}} \tag{6}$$

$$H_{i,j} \leftarrow H_{i,j} \frac{(W^T V)_{ij}}{(W^T W H)_{ij}} \tag{7}$$

Oleh karena itu, dalam penelitian ini, dalam salah satu tahap *preprocessing* data curah hujan harian dengan *Singular Spectrum Analysis* (SSA) adalah proses *dekomposisi*, dengan pendekatan alternatif *Non-negative Matrix Factorization* (NMF), agar output data preprocessing curah hujan tidak negative.

## 2.5. Processing

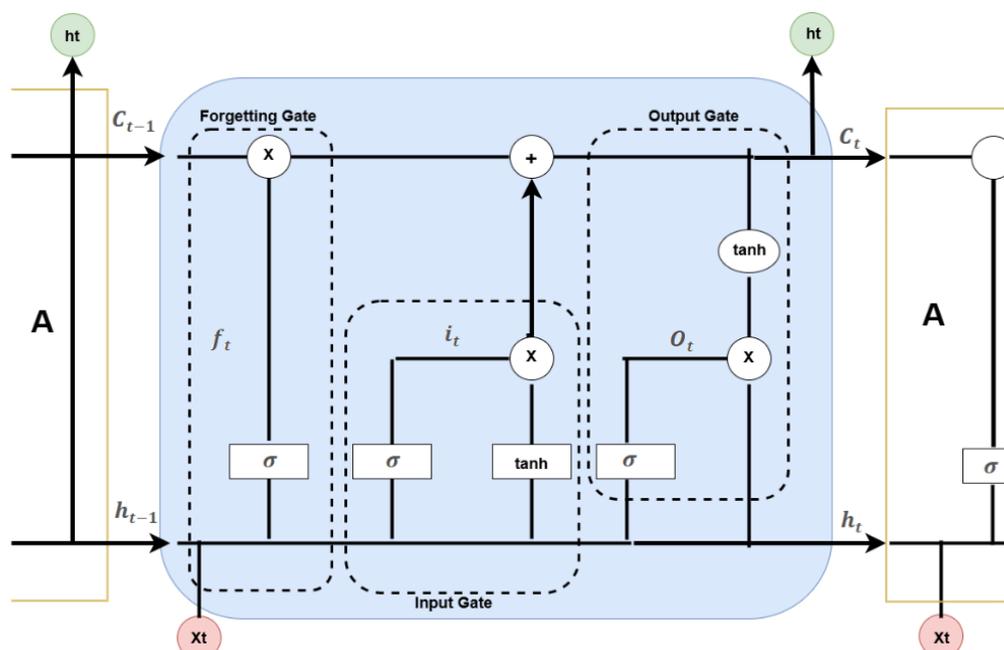
Hasil rekonstruksi dari proses *Singular Spectrum Analysis* (SSA) dimanfaatkan sebagai data latih dan validasi dalam pelatihan dua model hybrid, yakni SSA-LSSVR dan SSA-LSTM. Tujuan pelatihan ini adalah untuk membantu model mengenali pola dan hubungan dalam data curah hujan. Model LSSVR bekerja dengan membentuk garis regresi terbaik guna meminimalkan kesalahan prediksi, sedangkan LSTM, yang dirancang khusus untuk data berurutan seperti deret waktu, unggul dalam menyimpan informasi jangka panjang, sehingga mampu menangkap pola temporal dengan baik [24].

### 2.5.1. Pemodelan Dengan LSSVR

Konsep LSSVR (*Least Squares Support Vector Regression*) pertama kali diperkenalkan oleh Suykens dan Vandewalle [25] dan dikenal lebih efisien secara komputasi dibandingkan SVM klasik karena menggunakan pendekatan *least squares*. Pemilihan model dengan LSSVR dikarenakan dapat memecahkan masalah optimisasi dengan menyelesaikan satu set persamaan linear, sedangkan SVR biasa harus menyelesaikan masalah *Quadratic Programming* (QP) yang lebih kompleks dan intensif secara komputasi. LSSVR menawarkan solusi yang jauh lebih efisien dan cepat untuk dilatih dibandingkan SVR standar, sambil tetap mempertahankan daya prediksi dan kemampuan generalisasi yang tinggi. LSSVR sangat dipengaruhi oleh pemilihan parameter seperti  $\epsilon$ ,  $C$ , dan  $\gamma$ , yang menentukan akurasi prediksi. Sejumlah penelitian menyatakan bahwa pemilihan kernel Gaussian yang tepat dapat meningkatkan performa prediksi secara signifikan [26], [27], [28], [29] dan [30].

### 2.5.2. Pemodelan Dengan LSTM

Sementara itu, model LSTM, terbukti sangat andal dalam menambang karakteristik temporal data historis dan meningkatkan akurasi prediksi melalui kemampuannya memahami ketergantungan jangka panjang dalam data [24].



Gambar 3. unit dasar jaringan LSTM

Dengan memperkenalkan struktur memori dan menggabungkan output dari node sebelumnya pada setiap langkah waktu, LSTM dapat menambahkan atau menghapus informasi status sel melalui struktur gerbang (gambar 3).

### 2.5.3. Validasi Model

Langkah validasi model adalah fase krusial dalam memastikan kemampuan model yang telah dilatih untuk beroperasi secara efektif pada data baru yang sebelumnya belum pernah diperoleh. Proses ini bertujuan untuk menilai performa model sambil menghindari terjadinya overfitting, yaitu situasi di mana model terlalu terikat pada data pelatihan. Pada model LSSVR, validasi dilakukan dengan menguji kemampuan prediksi model terhadap data validasi, sedangkan pada model LSTM, validasi berfokus pada seberapa baik model dapat mempelajari pola urutan dari data yang baru.

### 2.5.4. Optimasi Hyperparameter

Setelah validasi, dilakukan proses optimasi hyperparameter menggunakan *RandomizedSearchCV* untuk LSSVR. Teknik ini bekerja dengan mencoba berbagai kombinasi parameter secara acak dari distribusi parameter yang telah ditentukan, untuk menemukan kombinasi yang paling optimal. Parameter yang dioptimasi mencakup  $C$  (*regularisasi*),  $\gamma$  (pengaruh titik data dalam kernel RBF), serta jenis kernel (RBF, *Linear*, atau *Polynomial*). Evaluasi menggunakan metrik MSE yang dibalik menjadi negatif untuk memaksimalkan skor model.

Sementara itu, *tuning hyperparameter* pada LSTM melibatkan pencarian nilai terbaik untuk parameter seperti jumlah unit *neuron*, *learning rate*, jumlah lapisan, *dropout rate*, dan *batch size*. Karena *hyperparameter* ini tidak bisa dipelajari otomatis saat pelatihan, penyesuaian manual sangat penting guna meningkatkan performa model.

Pendekatan sistematis seperti *RandomizedSearchCV* untuk LSSVR lebih efisien secara komputasi dibandingkan dengan *GridSearchCV* yang mencoba setiap kemungkinan kombinasi dari grid parameter yang diberikan. *RandomizedSearchCV* hanya mengambil sampel sejumlah kombinasi parameter secara acak dari distribusi yang ditentukan, membuat jauh lebih cepat dan cepat menemukan solusi yang sama baiknya atau bahkan lebih baik dalam waktu yang jauh lebih singkat. Pencarian *hyperparameter* terstruktur untuk LSTM memiliki arsitektur yang jauh lebih kompleks daripada LSSVR, dan performanya sangat sensitif terhadap berbagai hyperparameter. Percobaan dengan mengubah satu parameter dapat secara drastis memengaruhi kinerja optimal dari parameter lainnya. Sehingga dengan pendekatan optimasi *hyperparameter* dipilih untuk meninggalkan metode manual yang subjektif dan tidak efisien. Model yang dihasilkan berasal dari proses optimisasi yang metodis dan dapat dipertanggungjawabkan secara ilmiah.

### 2.5.5. Evaluasi Kinerja Model

Setelah proses tuning, dilakukan evaluasi kinerja model LSSVR dan LSTM menggunakan berbagai metrik seperti MSE, RMSE, MAE, dan NSE. MSE menghitung rata-rata kuadrat perbedaan antara yang diprediksi dan nilai yang sebenarnya [31], RMSE menghasilkan nilai yang mudah dipahami karena satuan yang digunakan identik dengan data yang asli, sedangkan MAE menghitung perbedaan rata-rata absolut nilai yang diprediksi dan nilai sebenarnya. [32], NSE mengindikasikan tingkat keakuratan prediksi model bila dibandingkan dengan rata-rata dari data yang sebenarnya [33].

## 2.6. Output

Kinerja setiap model hibrid dianalisis untuk mengidentifikasi model paling efektif yang dapat diaplikasikan dalam estimasi curah hujan harian di area Kota Semarang. Tahap akhir adalah visualisasi spasial hasil prakiraan menggunakan ArcGIS, dengan metode *interpolasi Inverse Distance Weighted*

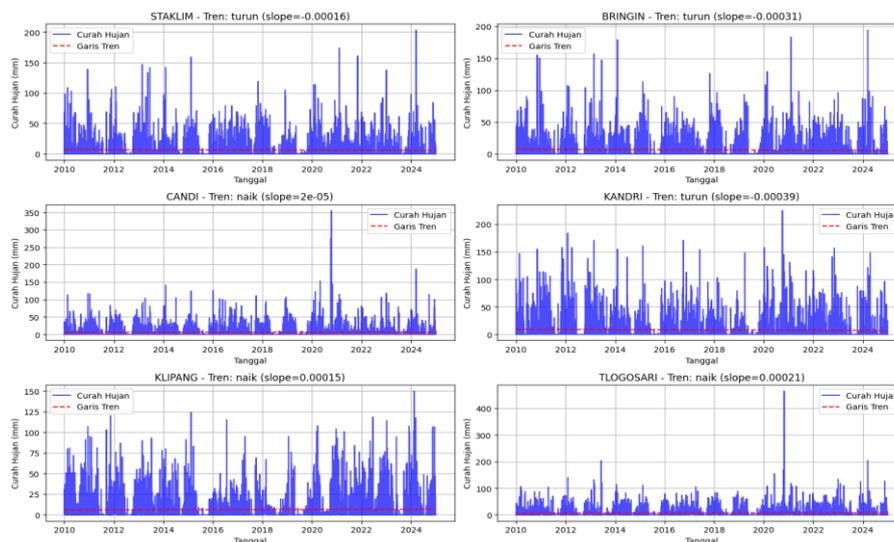
(IDW) berdasarkan data titik curah hujan di Kota Semarang. IDW menunjukkan tingkat ketepatan dan kecermatan yang lebih unggul dibandingkan dengan teknik lain seperti Kernel Density Estimation (KDE), sehingga menjadikannya pilihan yang ideal dalam menciptakan peta prakiraan curah hujan untuk berbagai situasi harian [34].

### 3. HASIL

#### 3.1. Variasi dan Tren Curah Hujan

Data Curah hujan yang berasal dari 6 (enam) lokasi pos pengamatan curah hujan di Kota Semarang diolah dalam bentuk grafik. Grafik ini untuk mengetahui variasinya dalam 15 tahun data mulai tahun 2010 – 2024 dan melihat tren perubahannya dalam periode tersebut, dapat dilihat pada Gambar 4.

Pada gambar 4, dapat dijelaskan bahwa seluruh pos hujan di 6 lokasi tersebut menghasilkan data curah hujan yang sangat variatif dari tidak ada hujan dalam pengukuran adalah 0 (nol) milimeter hingga 456 milimeter dalam satu hari yang dicatat di pos hujan Tlogosari. Tentunya dengan curah hujan setinggi itu berdampak pada potensi genangan atau banjir di wilayah tersebut. Garis tren dari masing-masing pos hujan untuk menunjukkan perbedaan perubahan curah hujannya dalam 15 tahun dengan nilai slope atau kemiringan garis tren naik atau turun perubahannya per hari. Stasiun Klimatologi Jawa Tengah dengan tren curah hujan turun 0,00016 mm/hari, pos hujan Bringin dengan tren curah hujan turun 0,00031 mm/hari dan Kandri dengan tren curah hujan turun 0,00039 mm/hari. Untuk pos hujan di Candi dengan tren curah hujan naik 0,00002 mm/hari, Klipang dengan tren curah hujan naik 0,00015 mm/hari dan Tlogosari dengan tren curah hujan naik 0,00021 mm/hari. Setiap perubahan naik dan turun curah hujan dalam satu hari meskipun kecil namun di masa depan sangat berpengaruh pada kecukupan air pada pertanian, perkebunan, air baku atau air minum, cadangan air, potensi bencana banjir, longsor, kekeringan dan kebakaran lahan/hutan.



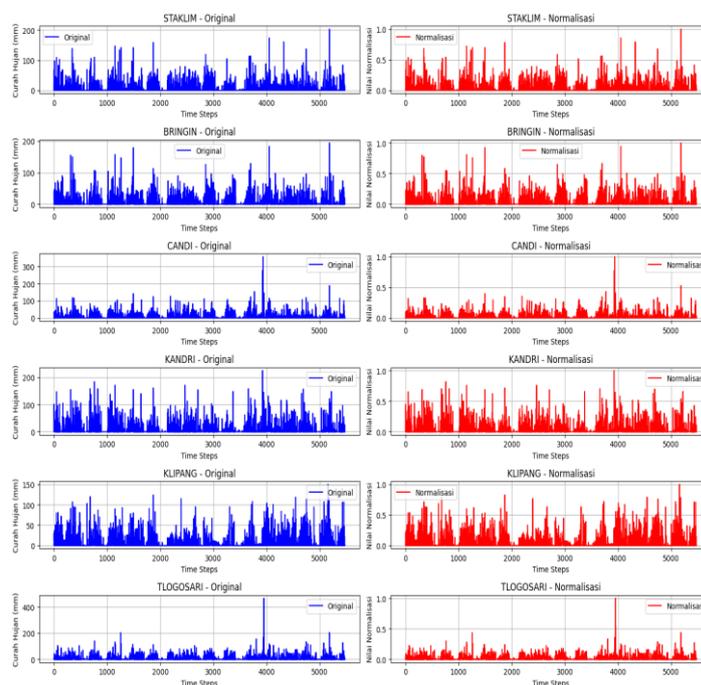
Gambar 4. Grafik curah hujan 6 lokasi pos hujan dan tren curah hujan periode 2010 – 2024 di Kota Semarang

#### 3.2. Normalisasi Data Hujan

Tahapan berikutnya pada gambar 6, untuk penyajian terpisah antara data asli dan hasil normalisasi dengan menggunakan *MinMaxScaler* (skala 0 hingga 1) dari library python yaitu *sklearn.preprocessing* dengan tetap menjaga pola tetapi merubah rentang nilainya, hal ini untuk memberikan gambaran utuh dari *raw data* hingga *preprocessing*. Total *time steps* ± 5500 hari mencakup data harian selama ± 15

tahun, disesuaikan dengan kebutuhan analisis jangka panjang. Proses ini bertujuan untuk memahami karakteristik spasial dan temporal curah hujan per stasiun, mengamati potensi *outlier* dan nilai ekstrem, menyiapkan data untuk model prediksi hybrid SSA-LSSVR dan SSA-LSTM yang sensitif terhadap skala dan memvalidasi keberhasilan proses normalisasi, agar bentuk pola temporal tetap terjaga.

Gambar 5 menyajikan visualisasi data curah hujan harian dari enam stasiun pengamatan di Kota Semarang, grafik kiri untuk curah hujan asli dan kanan adalah curah hujan yang sudah dinormalisasi. Stasiun Klimatologi Jawa Tengah pada data original menunjukkan banyak fluktuasi intensitas hujan dengan beberapa puncak tajam melebihi 150 mm/hari, dengan pola curah hujan tetap dipertahankan, namun skala disesuaikan dengan normalisasi. Pos hujan Bringin menunjukkan grafik curah hujan yang asli yang mengindikasikan volume hujan tinggi di sejumlah lokasi, sementara grafik normalisasi tetap mempertahankan ciri-ciri hujan yang ekstrem, tetapi disesuaikan dengan skala. Pos hujan Candi pada grafik original/asli menunjukkan curah hujan cenderung lebih rendah dibanding pos hujan lainnya, dengan satu puncak curah hujan signifikan sekitar >200 mm, sedangkan pada grafik normalisasi menunjukkan pola lonjakan besar tersebut tetap terlihat dengan nilai mendekati 1. Pos hujan Kandri pada grafik original data menunjukkan distribusi intensitas hujan cukup merata dengan frekuensi kejadian hujan tinggi cukup sering terjadi, dengan grafik normalisasi menunjukkan profil serupa dengan keterjagaan pola temporal. Pos hujan Klipang pada grafik original menunjukkan data yang lebih stabil, dengan intensitas maksimum yang tidak setinggi pos hujan lainnya, sedangkan pada grafik normalisasi menunjukkan distribusi normal terlihat sangat baik. Pos hujan Tlogosari grafik original cukup menarik karena menunjukkan fluktuasi tinggi untuk kejadian puncak hujan yang sangat ekstrem, sedangkan grafik normalisasi menunjukkan lonjakan ekstrem tetap terlihat meskipun telah dinormalisasi, menunjukkan karakteristik khas dari pos hujan ini pada periode 15 tahun tersebut.



Gambar 5. Grafik visualisasi data curah hujan harian antara data original/asli dan data hasil normalisasi.

### 3.3. Rekonstruksi SSA dan NMF

Langkah selanjutnya adalah proses rekonstruksi dengan SSA (*Singular Spectrum Analysis*) digunakan untuk mendekomposisi data waktu (*time series*) menjadi beberapa komponen utama, yang kemudian diproses *dekomposisi* dengan NMF (*Non-Negative Matrix Factorization*). Dengan NMF

digunakan untuk mendapatkan representasi data yang lebih terstruktur dan dapat diinterpretasi mendekati nilai normalisasi asli, mengekstrak tren dan pola utama. NMF bekerja baik pada data dengan banyak nilai nol dan dapat mengidentifikasi pola musiman atau lokasi dengan curah hujan tinggi. Hasil *dekomposisi* NMF selalu *non-negatif*, sehingga lebih interpretable untuk data curah hujan, yang sangat membantu dalam meningkatkan akurasi prediksi pada model selanjutnya. SSA + NMF bekerja untuk memperhalus data asli dan mengurangi variasi yang tidak relevan. Tujuan utama pada proses rekonstruksi ini adalah menilai efektivitas *dekomposisi* SSA + NMF dalam menangkap komponen utama dari data curah hujan, menganalisis rekonstruksi pola temporal yang mendekati data asli. Rekonstruksi ini sangat penting untuk prediksi berbasis model, membandingkan perbedaan antara data asli dan hasil rekonstruksi, yang menunjukkan kemampuan SSA + NMF dalam menghilangkan noise atau fluktuasi yang tidak relevan.

SSA akan melakukan *dekomposisi matriks Hankel* untuk mengekstrak pola utama dari data, dengan setting ukuran jendela (*window size*) adalah 90, digunakan untuk menangkap pola temporal dari data deret waktu selama 90 hari (tri wulan atau tiga bulan), karena semakin besar nilainya, semakin luas pola yang bisa diidentifikasi. Setting jumlah komponen latennya adalah 20, karena dalam *Non-negative Matrix Factorization* (NMF) digunakan untuk menentukan berapa banyak faktor yang akan dipertahankan untuk merepresentasikan data. Proses berikutnya data dibentuk dalam array 1 (satu) dimensi agar dapat diolah oleh SSA yang nantinya akan menghasilkan data curah hujan. Hasil olahan SSA digunakan sebagai komponen utama data yang selanjutnya untuk pengolahan menggunakan NMF.

Data yang dihasilkan dari *dekomposisi matriks Hankel* menggunakan metode SSA kemudian digunakan dalam proses inisialisasi NMF. Pada tahap ini, objek NMF disiapkan untuk melakukan *dekomposisi matriks* dengan inisialisasi acak, menggunakan random seed 42 agar hasil tetap konsisten saat dijalankan ulang. Jumlah iterasi maksimum ditetapkan sebanyak 500 untuk memperoleh faktor-faktor yang paling dominan. Selanjutnya dilakukan faktorisasi data curah hujan dengan matriks basis  $W$ , dimana  $W$  adalah *matriks non-negatif* yang berisi representasi fitur utama dari data. Menyimpan matrik  $H$  yang merepresentasikan kontribusi masing-masing fitur terhadap setiap faktor, dimana  $H$  adalah matriks non-negatif. Langkah berikutnya adalah rekonstruksi data dengan melakukan perkalian matriks  $W$  dan  $H$  untuk merekonstruksi data awal curah hujan, meskipun hasil data curah hujan rekonstruksi tidak identik dengan data awal curah hujan, ia tetap mempertahankan pola utama yang diekstrak oleh NMF. Data curah hujan diperoleh melalui penggabungan metode SSA untuk ekstraksi fitur dari data deret waktu dan NMF untuk faktorisasi guna memperoleh representasi laten. Hasil akhir dari proses SSA+NMF ini kemudian digunakan untuk mengidentifikasi tren dan pola tersembunyi, yang selanjutnya dimanfaatkan dalam prediksi curah hujan menggunakan metode LSSVR dan LSTM.

Pada gambar 6, grafik kiri menunjukkan data asli/original yang sudah di normalisasi, kemudian pada grafik sebelah kanan adalah hasil rekonstruksi data menggunakan SSA + NMF, yang berfungsi untuk menangkap pola temporal yang relevan dengan data asli.

Stasiun Klimatologi Jawa Tengah pada grafik original menunjukkan fluktuasi curah hujan yang cukup signifikan dengan puncak intensitas yang tercatat cukup tinggi. Pada grafik rekonstruksi SSA + NMF menunjukkan proses rekonstruksi mencoba untuk menangkap pola yang lebih halus dari data asli, meskipun ada beberapa perbedaan kecil pada puncak, rekonstruksi cukup baik dalam mengikuti tren secara umum.

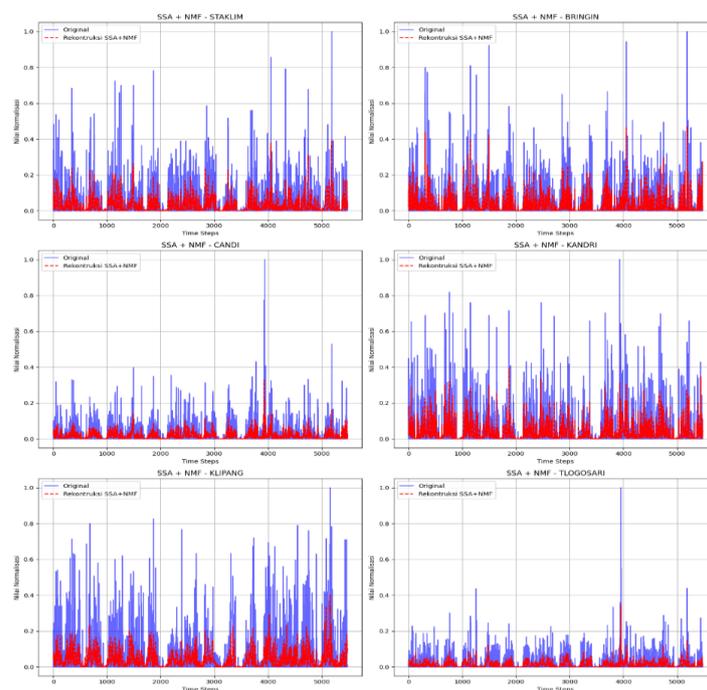
Pos hujan Bringin pada grafik asli menunjukkan pola hujan yang cenderung lebih stabil, dengan kejadian curah hujan sedang hingga tinggi. Pada grafik rekonstruksi SSA + NMF menunjukkan hasil rekonstruksi mencerminkan tren yang lebih halus dan mendekati data asli, meskipun variasi kecil tetap ada, secara keseluruhan, SSA + NMF mampu menangkap pola hujan utama.

Pos hujan Kandri pada grafik asli menunjukkan data curah hujan memiliki distribusi intensitas yang beragam, termasuk beberapa kejadian puncak hujan yang cukup tinggi, sedangkan pada grafik

rekonstruksi SSA + NMF menunjukkan data terlihat cukup baik dalam mengikuti pola fluktuasi. Terdapat beberapa variasi pada puncak ekstrem, namun SSA + NMF cukup efektif dalam menangkap komponen utama pola data hujan.

Pos hujan Klipang pada grafik original menunjukkan beberapa periode curah hujan tinggi yang sangat signifikan, sedangkan pada grafik rekonstruksi SSA + NMF menunjukkan pola yang lebih terhalus, tetapi tetap berhasil menangkap kejadian hujan ekstrem meskipun sedikit terdistorsi.

Pos hujan Tlogosari pada grafik original menunjukkan fluktuasi hujan yang sangat tajam, dengan puncak curah hujan yang luar biasa tinggi pada beberapa titik. Pada grafik rekonstruksi SSA + NMF menunjukkan rekonstruksi berhasil mempertahankan karakteristik tren utama data, namun ada beberapa perbedaan kecil dalam intensitas puncak, namun pola temporal utama tetap terjaga, menunjukkan keberhasilan dekomposisi SSA + NMF.



Gambar 6. Grafik visualisasi SSA + NMF antara data Asli dan data hasil rekonstruksi.

### 3.4. Pembelajaran Dengan Machine Learning (LSSVR) dan Deep Learning (LSTM)

Data curah hujan yang telah diolah menggunakan SSA NMF yang sudah halus dan bersih dari gangguan kemudian dimanfaatkan sebagai sumber untuk pengolahan data uji dan data latih dengan proporsi 80:20, di mana 80% digunakan sebagai data latih dan 20% sebagai data uji yang diterapkan dalam proses pembelajaran mesin menggunakan LSSVR (*Least Square Support Vector Regression*) serta dalam proses pembelajaran *deep learning* dengan LSTM (*Long Short Term Memory*). Dari masing-masing proses dilakukan dengan tuning, agar dapat memperoleh analisa, pengolahan dan pembelajaran yang maksimal untuk memperoleh prediksi yang lebih baik.

Dalam penelitian ini tuning LSSVR (*Least Squares Support Vector Regression*) dengan metode *RandomizedSearchCV* melakukan pencarian parameter acak dengan mencoba 50 kombinasi parameter yang berbeda, untuk setiap kombinasi parameter, dilakukan validasi *5 fold crossvalidation* (data dibagi menjadi 5 bagian, 4 untuk training dan 1 untuk testing, diulang 5 kali). Sehingga total model yang dilatih adalah  $5 \text{ folds} \times 50 \text{ kombinasi}$  hasilnya adalah 250 model, hal ini dilakukan untuk memastikan model tidak overfitting dan stabil terhadap pembagian data. Hasil tuning LSSVR dari masing-masing pos hujan adalah sebagai berikut :

Tabel 3. Hasil tuning LSSVR dengan metode *RandomizedSearchCV* masing-masing pos hujan

| Pos Hujan | PARAMETER |       |        |       |
|-----------|-----------|-------|--------|-------|
|           | C         | Gamma | Kernel | MSE   |
| Staklim   | 47.659    | 0.096 | Linear | 0.002 |
| Bringin   | 656.913   | 0.751 | Linear | 0.002 |
| Candi     | 4.043     | 0.347 | Linear | 0.005 |
| Kandri    | 47.659    | 0.096 | Linear | 0.004 |
| Klipang   | 237.984   | 0.004 | RBF    | 0.002 |
| Tlogosari | 237.984   | 0.004 | RBF    | 0.003 |

Pada Tabel 3, untuk Stasiun Klimatologi Jawa Tengah model dengan kernel linear mampu memberikan performa sangat baik, terlihat dari MSE yang rendah (0.002). Nilai C yang moderat dan gamma yang kecil menunjukkan regulasi dan sensitivitas model yang seimbang. Pos Hujan Bringin juga optimal dengan *kernel linear*, namun nilai C dan gamma cukup besar, menunjukkan bahwa model cenderung lebih kompleks dan sensitif terhadap perubahan data meski demikian, MSE tetap rendah, menandakan *overfitting* tidak terjadi. Pos Hujan Candi memiliki performa lebih tinggi dibanding pos lainnya (MSE tertinggi) namun dalam performa evaluasi yang rendah. Nilai C yang rendah menunjukkan model lebih kuat terhadap regularisasi. Pos Hujan Kandri mirip dengan Stasiun Klimatologi Jawa Tengah dalam hal parameter, tetapi MSE sedikit lebih tinggi namun tetap dalam performa yang tinggi. Pos Hujan Klipang bekerja optimal dengan *kernel non-linear* (RBF), menunjukkan bahwa hubungan antara fitur dan curah hujan di sini tidak linear. MSE yang sangat rendah menunjukkan model berhasil memetakan hubungan tersebut dengan baik. Sama seperti Pos Hujan Klipang, Pos Hujan Tlogosari menggunakan kernel RBF dengan parameter identik. MSE sedikit lebih tinggi, bisa jadi karena data yang lebih kompleks atau noise di pos ini.

Memperoleh performa optimal dari model *Long Short-Term Memory* (LSTM), dilakukan proses tuning hyperparameter secara sistematis menggunakan *Keras Tuner* dengan pendekatan *RandomSearch*. Hal ini memungkinkan eksplorasi cepat ruang *hyperparameter* secara acak, lebih efisien dibanding *Grid Search* pada kasus dengan banyak kombinasi, tanpa harus mencoba seluruh kemungkinan kombinasi, menjadikannya pilihan yang tepat untuk tuning model LSTM pada dataset. Model LSTM yang diterapkan dalam studi ini terdiri dari dua lapisan LSTM yang disusun secara bertingkat.

Tujuan dari tuning ini adalah untuk mendapatkan kombinasi *hyperparameter* terbaik yang mampu meminimalkan validation loss. Maksimal 10 kombinasi *hyperparameter* akan dicoba, setiap kombinasi hanya diuji sekali untuk efisiensi, 50 iterasi (*epochs*) maksimum untuk setiap kombinasi, dengan 32 data tiap iterasi. Dilanjutkan, pemisahan validasi 20% dari data pelatihan digunakan untuk tujuan validasi, dan digunakan *early stopping* untuk menghentikan proses pelatihan jika tidak terdapat peningkatan selama 5 epoch, guna menghindari *overfitting*. Setelah proses pencarian selesai, model dengan performa terbaik diambil untuk kemudian dilatih ulang dengan dataset yang sama. Konfigurasi optimal dengan *epochs* hingga 100 untuk hasil akhir yang lebih stabil, *earlystopping* kini dengan *patience* = 10 agar memberi ruang untuk generalisasi yang lebih baik untuk mencegah *overfitting* serta mempercepat proses pelatihan. Sehingga model memiliki kemampuan generalisasi yang baik dalam memprediksi data curah hujan hasil dekomposisi SSA-NMF.

Tabel 4 menyajikan hasil *tuning hyperparameter* model LSTM untuk masing-masing pos hujan menggunakan metode *Random Search* yang diimplementasikan melalui *Keras Tuner*. Parameter yang dioptimalkan mencakup jumlah unit pada dua lapisan LSTM, nilai learning rate, serta nilai *Mean Squared Error* (MSE) terendah yang dicapai selama proses pelatihan. Untuk seluruh pos hujan di Kota Semarang, kombinasi terbaik diperoleh dengan menggunakan 64 unit pada masing-masing lapisan LSTM. Hal ini menunjukkan bahwa arsitektur model dengan kapasitas memori sebesar 64 unit pada

setiap lapisan mampu menangkap pola temporal dari data deret waktu curah hujan dengan baik di seluruh lokasi. Konsistensi pemilihan jumlah unit ini mengindikasikan bahwa kompleksitas jaringan sebesar ini cukup stabil dan representatif untuk data yang digunakan. Nilai learning rate terbaik yang dihasilkan dari tuning adalah sebesar 0.01 untuk semua pos hujan. Learning rate ini menunjukkan kecepatan model dalam memperbarui bobot selama proses pelatihan. Nilai ini dianggap optimal karena cukup cepat untuk konvergen tanpa menyebabkan lonjakan atau ketidakstabilan selama proses optimasi. Kolom terakhir menunjukkan nilai MSE terendah yang dicapai oleh model setelah tuning. Nilai MSE ini dipakai sebagai pengukuran utama karena menunjukkan rata-rata kuadrat perbedaan antara angka yang sebenarnya dan angka yang diprakirakan untuk curah hujan. Hasil tuning menunjukkan bahwa model memberikan performa terbaik di Stasiun Klimatologi Jawa Tengah (MSE = 0.000098928), diikuti oleh Candi (0.000059168) dan Tlogosari (0.000034531). Sementara itu, performa lebih rendah terjadi di pos Bringin (MSE = 0.00029984) dan Kandri (0.00025740). Variasi nilai MSE ini mengindikasikan adanya perbedaan karakteristik pola curah hujan antar lokasi, yang mungkin dipengaruhi oleh faktor geografis atau dinamika cuaca dan iklim mikro setempat.

Tabel 4. Hasil tuning LSTM dengan metode *hyperparameter* masing-masing pos hujan

| Pos Hujan | PARAMETER        |                    |               |              |
|-----------|------------------|--------------------|---------------|--------------|
|           | Jumlah Unit LSTM | Jumlah Unit LSTM_1 | Learning Rate | MSE Terendah |
| Staklim   | 64               | 64                 | 0.01          | 0.000098928  |
| Bringin   | 64               | 64                 | 0.01          | 0.00029984   |
| Candi     | 64               | 64                 | 0.01          | 0.000059168  |
| Kandri    | 64               | 64                 | 0.01          | 0.00025740   |
| Klipang   | 64               | 64                 | 0.01          | 0.00013829   |
| Tlogosari | 64               | 64                 | 0.01          | 0.000034531  |

### 3.5. Training dan Validation Loss

*Training loss* berfungsi untuk mengukur seberapa baik model LSTM belajar dari data latih (*training set*) dengan tujuan untuk menurunkan nilai training loss sehingga model semakin memahami pola dari data latihnya. *Validation loss* untuk mengukur seberapa baik model berperformansi pada data validasi (*validation set*) yang tidak digunakan untuk pelatihan langsung dengan tujuan untuk memastikan model tidak hanya bekerja baik pada data latih, tapi juga bisa generalisasi ke data baru tanpa kehilangan akurasi. Jika *validation loss* mulai meningkat sementara *training loss* tetap turun, berarti model mengalami *overfitting* (atau istilahnya model terlalu ‘hafal’ data latih dan kurang mampu menangani data baru). Berikut ini adalah hasil proses dalam pelatihan model LSTM di 6 pos hujan di Kota Semarang dengan dua metrik utama yang di evaluasi yaitu *training loss* dan *validation loss* pada gambar 7.

Gambar 7 menjelaskan hasil training dan *validation loss* dari 6 pos hujan di Kota Semarang, dengan hasil di Stasiun Klimatologi Jawa Tengah pola loss terlihat cukup stabil setelah epoch ke-10. Awal pelatihan menunjukkan fluktuasi signifikan, tapi setelah itu training dan *validation loss* saling mendekati dan menurun bersama, sehingga model menunjukkan konvergensi yang baik dan tidak *overfitting*, hal ini sejalan dengan nilai MSE rendah di tabel 4 yaitu 0.000098928.

Pos Hujan Bringin fluktuasi *validation loss* cukup tinggi dan tidak stabil, bahkan kadang lebih rendah dari training loss dengan durasi pelatihan sampai 45 *epoch*. Ada indikasi *overfitting* ringan karena training loss terus turun, tetapi *validation loss* cenderung tetap dan tidak membaik secara

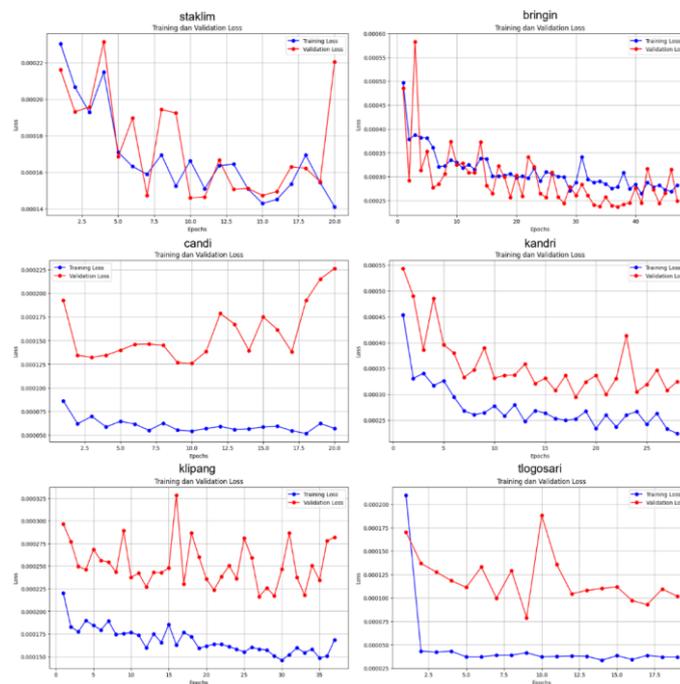
signifikan, hal ini konsisten dengan MSE yang lebih tinggi dibanding pos lainnya dengan nilai MSE adalah 0.00029984.

Pos hujan Candi *training loss* cukup stabil dan rendah, tetapi *validation loss* cenderung tinggi dan berfluktuasi kuat, dengan durasi pelatihan 20 *epoch*, *overfitting* cukup jelas terjadi karena model belajar terlalu baik pada data latih, tetapi gagal generalisasi. Walau begitu, MSE masih tergolong rendah yaitu 0.000059168, menandakan ada potensi peningkatan dengan teknik regularisasi atau augmentasi.

Pos hujan Kandri *training loss* turun stabil, *validation loss* fluktuatif, tetapi cenderung stagnan, dengan durasi pelatihan 27 *epoch*, disimpulkan sama seperti Bringin dan Candi, ada indikasi *overfitting* sedang, terlihat dari gap antara *training* dan *validation loss* yang tidak mengecil.

Pos hujan Klipang *training loss* menurun stabil, *validation loss* sangat fluktuatif dan tidak menunjukkan pola menurun yang konsisten dengan durasi pelatihan 36 *epoch*, *overfitting* paling tinggi dari semua lokasi pos hujan di Kota Semarang, *training loss* sangat rendah tetapi *validation loss* tidak menurun secara signifikan, bahkan cenderung naik-turun tajam, namun nilai MSE masih tergolong rendah yaitu 0.00013829.

Pos hujan Tlogosari hasil *training* dan *validation loss* menurun dengan pola yang cukup harmonis, meski ada sedikit spike di pertengahan dengan durasi pelatihan 19 *epoch*, model berhasil belajar dengan baik, hampir tidak ada *overfitting*. Hal ini juga selaras dengan MSE rendah (0.000034531), menandakan performa generalisasi yang baik. Secara keseluruhan pos hujan dengan hasil terbaik adalah Tlogosari dan Staklim (stabil & minim *overfitting*) dan pos hujan yang perlu perhatian lebih adalah Klipang dan Candi (indikasi *overfitting*).



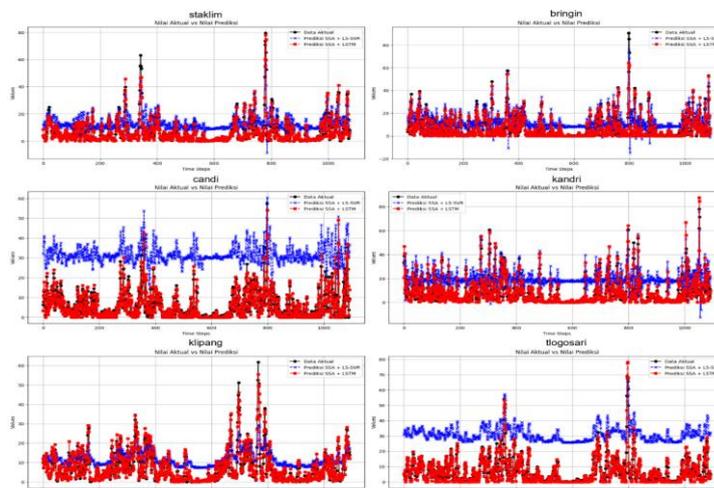
Gambar 7. Grafik training dan validation loss masing-masing pos hujan di Kota Semarang

### 3.6. Perbandingan Hasil Nilai aktual Dengan Nilai Prediksi

Berdasarkan hasil visualisasi prediksi yang ditampilkan pada Gambar 8, dapat diamati bahwa performa model hybrid SSA + LSTM secara umum menunjukkan hasil yang lebih akurat dalam mengikuti pola data curah hujan aktual dibandingkan dengan model SSA + LSSVR. Garis merah yang merepresentasikan hasil prediksi SSA + LSTM tampak lebih mampu mengikuti dinamika fluktuasi data aktual (garis hitam), termasuk dalam menangkap lonjakan-lonjakan ekstrem curah hujan. Hal ini

menunjukkan bahwa model LSTM, dengan kemampuannya dalam mempelajari dependensi jangka panjang dari data deret waktu, berhasil mengadaptasi variasi pola curah hujan yang kompleks dan tidak linier.

Sebaliknya, garis biru yang merepresentasikan prediksi SSA + LSSVR menunjukkan hasil prediksi yang cenderung lebih halus dan konservatif. Meskipun prediksi model ini stabil dan tidak menghasilkan nilai ekstrem yang fluktuatif, karakteristik tersebut justru menjadi kelemahan ketika model dihadapkan pada data aktual dengan lonjakan-lonjakan curah hujan yang tinggi. SSA + LSSVR kurang mampu mengikuti puncak-puncak curah hujan yang signifikan, sehingga menghasilkan prediksi yang seringkali berada di bawah nilai aktual.



Gambar 8. Grafik perbandingan nilai aktual dengan nilai prediksi dengan metode hybrid SSA+LSSVR dan SSA+LSTM pada masing-masing pos hujan di Kota Semarang

Seluruh pos hujan yang diamati yaitu Staklim, Bringin, Candi, Kandri, Klipang, dan Tlogosari polanya terlihat konsisten. Model SSA + LSTM mampu menghasilkan prediksi yang lebih mendekati data aktual di berbagai kondisi, baik saat curah hujan rendah maupun tinggi. Hal ini mengindikasikan bahwa integrasi SSA sebagai metode dekomposisi dan pembersih derau data, dikombinasikan dengan kemampuan LSTM dalam menangani data sekuensial dan temporal, memberikan keunggulan dalam konteks prakiraan curah hujan harian.

Secara keseluruhan, dapat disimpulkan bahwa model hybrid SSA + LSTM lebih unggul dibandingkan SSA + LSSVR, terutama dalam aspek presisi dan sensitivitas terhadap fluktuasi data curah hujan aktual. Hasil ini mendukung penggunaan pendekatan deep learning berbasis LSTM sebagai metode utama dalam pengembangan sistem prakiraan curah hujan berbasis data time series terdekomposisi, khususnya dalam upaya mitigasi risiko bencana hidrometeorologi di wilayah penelitian yaitu di Kota Semarang.

### 3.7. Evaluasi Hasil Prediksi Model Hybrid

Sesuai dengan penilaian yang terdapat di Tabel 5 dan Tabel 6, kinerja model hibrid SSA+LSTM umumnya menunjukkan performa yang lebih unggul daripada model SSA+LSSVR di semua stasiun hujan (pos hujan) yang telah dikaji, yaitu Staklim (Stasiun Klimatologi Jawa Tengah), Bringin, Candi, Kandri, Klipang, dan Tlogosari. Evaluasi dilakukan menggunakan empat metrik: *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), dan *Nash-Sutcliffe Efficiency* (NSE).

Akurasi model berdasarkan MSE dan RMSE pada model SSA+LSTM menunjukkan nilai yang jauh lebih rendah dibandingkan dengan SSA+LSSVR. Sebagai contoh di Pos Hujan Candi, model

SSA+LSSVR menghasilkan MSE sebesar 677.80 dan RMSE 26.03, sedangkan model SSA+LSTM hanya sebesar 8.73 untuk MSE dan 2.96 untuk RMSE. Kemudian di Pos Hujan Staklim (Stasiun Klimatologi Jawa Tengah), SSA+LSTM menghasilkan MSE 9.05 dan RMSE 3.01, jauh lebih baik dibanding SSA+LSSVR yang menghasilkan MSE 69.27 dan RMSE 8.35. Semua lokasi pos hujan di Kota Semarang menunjukkan bahwa prediksi dari SSA+LSTM lebih mendekati nilai aktual dibandingkan dengan SSA+LSSVR.

Konsistensi dan stabilitas kesalahan prediksi berdasarkan nilai evaluasi MAE, model SSA+LSTM juga lebih rendah, menunjukkan bahwa kesalahan absolut rata-rata dari model ini hanya menghasilkan MAE sebesar 1.76 dibandingkan dengan 6.13 pada SSA+LSSVR. Kemudian di pos hujan Tlogosari, SSA+LSTM memiliki MAE sebesar 1.99 jauh lebih kecil dibandingkan SSA+LSSVR sebesar 24.10. Demikian juga untuk pos hujan lainnya di Kota Semarang prediksi dari SSA+LSTM lebih mendekati nilai aktual dibandingkan dengan SSA+LSSVR.

Kinerja Efisiensi Model dengan menggunakan NSE, merupakan indikator penting untuk menilai efisiensi model dalam mereplikasi variasi data observasi. Model SSA+LSTM memperlihatkan nilai NSE yang sangat baik di semua pos hujan dengan nilai > 0.8 bahkan mencapai 0.90 di Bringin dan 0.89 di Klipang. Sebaliknya, model SSA+LSSVR menghasilkan nilai NSE negatif di beberapa lokasi seperti: Candi dengan NSE -12.85, Tlogosari dengan NSE -7.53 dan Kandri dengan NSE -0.69. NSE negatif menandakan bahwa model lebih buruk daripada sekadar menggunakan rata-rata data aktual sebagai prediksi, hal ini menunjukkan bahwa model SSA+LSSVR tidak mampu menangkap pola data dengan baik di lokasi-lokasi tersebut.

Model hybrid SSA+LSTM secara signifikan lebih unggul dibandingkan SSA+LSSVR dalam memprediksi curah hujan harian di semua pos hujan yang dianalisis. Hal ini dibuktikan melalui nilai MSE, RMSE, dan MAE yang lebih rendah serta NSE yang lebih tinggi dan positif di seluruh lokasi. Dengan demikian, model SSA+LSTM lebih direkomendasikan untuk digunakan dalam prakiraan curah hujan harian sebagai bagian dari sistem peringatan dini terhadap potensi bencana hidrometeorologi.

Tabel 5. Hasil evaluasi prediksi harian menggunakan model hybrid SSA + LSSVR

| Pos Hujan | MODEL HYBRID SSA + LSSVR |       |       |        |
|-----------|--------------------------|-------|-------|--------|
|           | MSE                      | RMSE  | MAE   | NSE    |
| Staklim   | 69.267                   | 8.35  | 7.52  | 0.15   |
| Bringin   | 63.45                    | 7.97  | 7.06  | 0.34   |
| Candi     | 677.80                   | 26.03 | 25.59 | -12.85 |
| Kandri    | 189.61                   | 13.77 | 12.83 | -0.69  |
| Klipang   | 50.29                    | 7.09  | 6.13  | 0.21   |
| Tlogosari | 602.85                   | 24.55 | 24.10 | -7.53  |

Tabel 6. Hasil evaluasi prediksi harian menggunakan model hybrid SSA + LSTM

| Pos Hujan | MODEL HYBRID SSA + LSTM |      |      |      |
|-----------|-------------------------|------|------|------|
|           | MSE                     | RMSE | MAE  | NSE  |
| Staklim   | 9.04                    | 3.01 | 1.76 | 0.89 |
| Bringin   | 9.34                    | 3.06 | 1.87 | 0.90 |
| Candi     | 8.73                    | 2.95 | 1.98 | 0.82 |
| Kandri    | 13.83                   | 3.72 | 2.30 | 0.88 |
| Klipang   | 7.17                    | 2.68 | 1.76 | 0.89 |
| Tlogosari | 9.03                    | 3.00 | 1.99 | 0.87 |

Rangkuman kinerja rata-rata dari model di semua lokasi pengamatan yang terdapat dalam Tabel 7 menunjukkan bahwa model SSA-LSTM menampilkan performa yang jauh lebih unggul secara signifikan dibandingkan dengan model SSA-LSSVR pada seluruh metrik evaluasi yang diterapkan. Model SSA-LSTM terus-menerus menghasilkan nilai kesalahan yang lebih rendah, seperti ditunjukkan oleh nilai RMSE sebesar 3.1 dan MAE sebesar 1.9 yang sangat kecil. Di sisi lain, model SSA-LSSVR mencatatkan kesalahan yang lebih besar dengan RMSE mencapai 14.6 dan MAE sebesar 13.9. Perbedaan performa yang paling penting terlihat pada metrik *Nash-Sutcliffe Efficiency* (NSE). Nilai NSE model SSA-LSTM mencapai 0.9, yang menunjukkan bahwa model tersebut sangat efisien dan dapat diandalkan dalam meniru data observasi (mendekati angka sempurna 1.0). Sebaliknya, nilai NSE negatif (-3.4) pada model SSA-LSSVR menunjukkan bahwa performa model tersebut jauh lebih buruk dibandingkan hanya menggunakan rata-rata data historis.

Secara keseluruhan, hasil ini secara jelas membuktikan bahwa arsitektur LSTM lebih unggul dalam menangkap pola non-linear pada data curah hujan yang telah diproses oleh SSA, menjadikannya model terbaik dalam penelitian ini.

Tabel 7. Rangkuman Rata-Rata Kinerja Model di Seluruh Lokasi Pengamatan

| Model     | Metrik | Rata-Rata (Average) |
|-----------|--------|---------------------|
| SSA-LSSVR | MSE    | 275.5               |
|           | RMSE   | 14.6                |
|           | MAE    | 13.9                |
|           | NSE    | -3.4                |
| SSA-LSTM  | MSE    | 9.5                 |
|           | RMSE   | 3.1                 |
|           | MAE    | 1.9                 |
|           | NSE    | 0.9                 |

### 3.8. Visualisasi Spasial Hasil Prediksi Model Hybrid Terbaik

Mendapatkan gambaran visualisasi spasial dengan menggunakan pemetaan ArcGIS dari hasil prediksi curah hujan harian selama 10 hari berdasarkan data curah hujan selama 15 tahun output model hybrid SSA-LSTM sebagai model yang terbaik dibandingkan dengan model SSA-LSSVR (tabel 8). Legend warna pada peta visualisasi tersebut adalah klasifikasi curah hujan harian berdasarkan standar BMKG. Setiap warna pada gambar peta spasial menunjukkan intensitas curah hujan selama akumulasi curah hujan satu hari (24 jam) dengan kriteria tidak hujan dengan curah hujan 0 – 0.5 mm/hari sampai intensitas tertinggi yaitu ekstrim dengan curah hujan > 150 mm/hari (tabel 9).

Penjelasan untuk masing-masing hasil visualisasi prediksi curah hujan harian spasialnya adalah pada gambar 9, dengan penjelasan adalah sebagai berikut, titik-titik warna merah dalam peta adalah pos hujan yang berada di Kota Semarang.

Hari ke-1, hari ke-2 dan ke-5 seluruh wilayah Kota Semarang diprediksi terjadi hujan dengan intensitas ringan dengan legend warna hijau muda. Pada hari ke-3 dan ke-4, diramalkan akan terjadi hujan dengan intensitas dari ringan sampai sedang, dengan warna hijau yang menunjukkan hujan ringan hampir mencakup seluruh area Kota Semarang. Sementara itu, warna kuning untuk menunjukkan hujan sedang diperkirakan akan muncul di daerah Kecamatan Candisari, sebagian wilayah Kecamatan Gajahmungkur, beberapa bagian di Kecamatan Semarang Selatan, serta sedikit wilayah di Kecamatan Banyumanik, Kecamatan Tembalang, dan Kecamatan Semarang Tengah.

Untuk hari ke-6, ke-7, dan ke-8, diperkirakan tingkat curah hujan akan berkisar antara tidak hujan sampai ringan. Dengan warna abu-abu menunjukkan tidak hujan, diprediksikan akan terjadi di area

Kecamatan Semarang Barat, Kecamatan Genuk, beberapa bagian Kecamatan Pedurungan, sebagian kecil dari Kecamatan Gayamsari, sebagian kecil dari Kecamatan Tugu, sebagian kecil Kecamatan Ngaliyan, sebagian kecil Kecamatan Semarang Utara, dan sebagian kecil Kecamatan Semarang Timur.

Terakhir, untuk prakiraan pada hari ke-9 dan ke-10, diperkirakan terdapat curah hujan dengan intensitas mulai dari tidak hujan hingga ringan. Warna abu-abu yang menunjukkan tidak hujan diprediksi akan meliputi sebagian area Kecamatan Pedurungan, sebagian wilayah Kecamatan Genuk, sebagian area Kecamatan Gayamsari, serta sedikit bagian dari Kecamatan Semarang Timur.

Tabel 8. Ouput prediksi curah hujan terbaik SSA - LSTM

| Har<br>i | Output Prediksi Curah Hujan SSA – LSTM (milimeter/hari) |         |       |        |         |           |
|----------|---|---------|-------|--------|---------|-----------|
|          | Staklim   | Bringin | Candi | Kandri | Klipang | Tlogosari |
| 1        | 8.99  | 8.56    | 10.40 | 10.98  | 12.42   | 4.78      |
| 2        | 8.14  | 8.56    | 14.82 | 10.99  | 13.08   | 4.47      |
| 3        | 6.19  | 6.51    | 20.09 | 10.92  | 14.02   | 4.02      |
| 4        | 3.15  | 5.06    | 21.41 | 12.56  | 13.75   | 2.83      |
| 5        | 1.14  | 5.05    | 18.94 | 14.12  | 12.40   | 1.08      |
| 6        | 0   | 5.69    | 15.67 | 14.86  | 11.08   | 0         |
| 7        | 0   | 6.70    | 12.61 | 14.76  | 10.15   | 0         |
| 8        | 0.23  | 7.51    | 10.02 | 14.77  | 9.30    | 0         |
| 9        | 1.30  | 8.07    | 8.60  | 15.32  | 8.18    | 0         |
| 10       | 2.75  | 8.26    | 8.07  | 15.99  | 7.02    | 0         |

Tabel 9. Klasifikasi curah hujan harian dan kriterianya (BMKG)

| Klasifikasi Curah Hujan | Intensitas Curah Hujan | Warna   |
|-------------------------|------------------------|---------|
| 0 – 0.5 mm/hari         | Tidak Hujan            | Abu-abu |
| 0.5 – 20 mm/hari        | Hujan Ringan           | Hijau   |
| 20 – 50 mm/hari         | Hujan Sedang           | Kuning  |
| 50 – 100 mm/hari        | Hujan Lebat            | Orange  |
| 100 – 150 mm/hari       | Hujan Sangat Lebat     | Merah   |
| > 150 mm/hari           | Ekstrim                | Magenta |

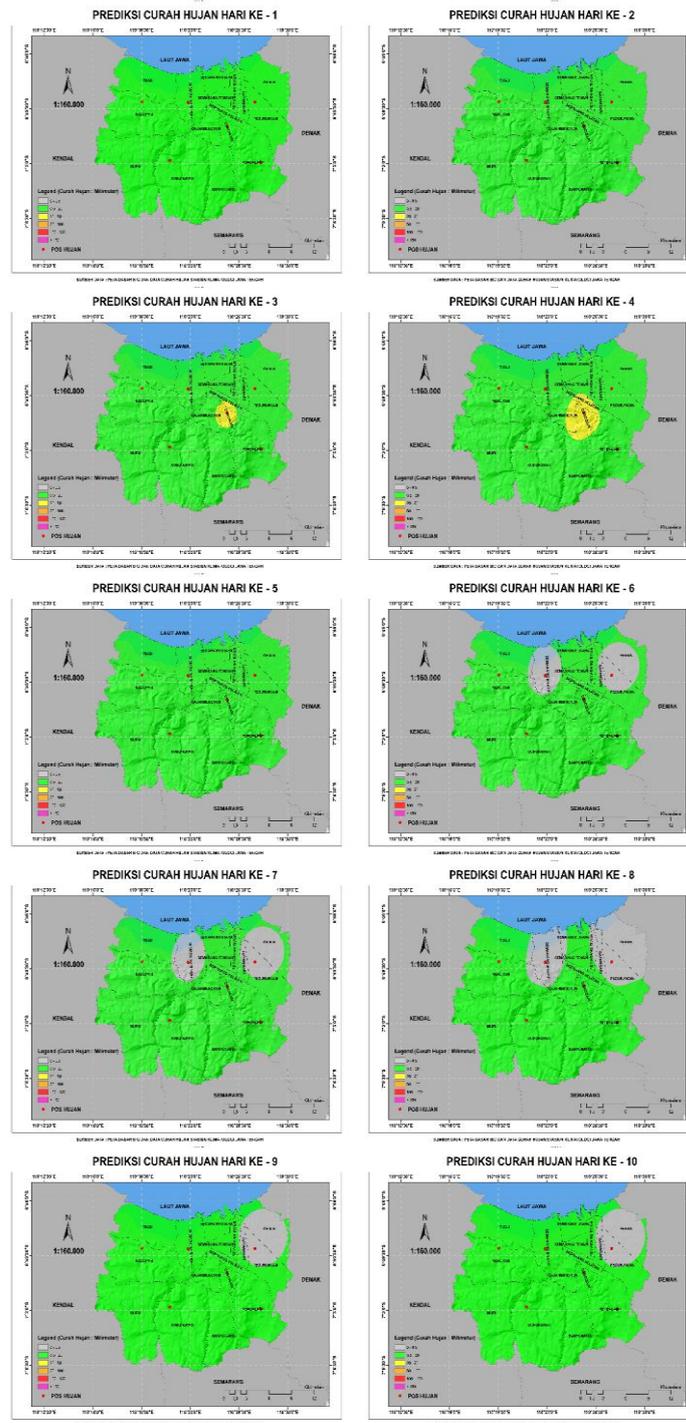
### 3.9. Analisa Error Outlier

Tabel 10 menyajikan ringkasan analisis *error outlier* rata-rata untuk kedua model hybrid (SSA-LSSVR dan SSA-LSTM) pada dataset 15 tahun (5.475 sampel harian, 80% latih, 20% uji).

Tabel 10. Ringkasan analisis *error outlier* rata-rata untuk LSSVR dan LSTM

| Model      | Total Sampel | Outliers | Non-Outliers | MAE All | RMSE All | MAE Non-Outliers | RMSE Non-Outliers | MAE Outliers | RMSE Outliers |
|------------|--------------|----------|--------------|---------|----------|------------------|-------------------|--------------|---------------|
| SSA-LSSVR  | 1089         | 56       | 1034         | 13.87   | 14.63    | 13.97            | 14.56             | 11.41        | 13.36         |
| SSA - LSTM | 1089         | 118      | 972          | 1.96    | 3.02     | 1.35             | 1.79              | 6.97         | 7.59          |

SSA-LSTM secara konsisten menunjukkan kesalahan yang lebih rendah dalam kondisi normal maupun bukan *outlier*, menunjukkan tingkat presisi yang tinggi. Dalam kasus *outlier*, SSA-LSTM tetap unggul dengan kesalahan sekitar 7 mm dibanding SSA-LSSVR yang mencapai kesalahan sekitar 12 mm. Namun, perbedaan kesalahan antara kondisi normal dan *outlier* pada SSA-LSTM lebih besar, yang menunjukkan kesulitan dalam memprediksi hujan ekstrem. Sementara itu, SSA-LSSVR memiliki kesalahan secara keseluruhan yang lebih tinggi, tetapi perbedaan antara kondisi normal dan *outlier* relatif lebih kecil. Hasil ini mendukung penggunaan SSA-LSTM untuk akurasi umum, meskipun perlu ada peningkatan khusus dalam memprediksi kejadian ekstrem.



Gambar 9. Visualisasi spasial hasil prediksi curah hujan harian dengan model hybrid terbaik yaitu SSA+LSTM

## 4. DISKUSI

Penelitian ini berawal dari keprihatinan mengenai semakin tingginya potensi bencana hidrometeorologi di kawasan tropis seperti Semarang, terutama disebabkan oleh pola hujan yang semakin sulit diprediksi dan cenderung tidak menentu. Maka diperlukan penelitian untuk menjawab tantangan ini dengan pendekatan yang lebih adaptif dan akurat. Kombinasi metode *Singular Spectrum Analysis* (SSA) dengan dua pendekatan yang berbeda *Least Squares Support Vector Regression* (LSSVR) dan *Long Short-Term Memory* (LSTM) merupakan strategi utama dalam penelitian ini. Tujuannya bukan hanya untuk menghasilkan prediksi curah hujan yang akurat, tetapi juga untuk membangun sistem peringatan dini berbasis data yang kuat.

### 4.1. Kinerja Model Hybrid SSA-LSSVR dan SSA-LSTM

Hasil penelitian menunjukkan bahwa model hybrid SSA-LSTM secara konsisten memberikan performa prediksi yang lebih baik dibandingkan SSA-LSSVR. Ini tampak dari nilai MSE, RMSE, MAE, dan terutama dari nilai NSE yang lebih tinggi pada hampir semua lokasi pos hujan. Misalnya, pada pos hujan Bringin, model SSA-LSTM mampu mencapai NSE serendah 0.9, menunjukkan sensitivitas tinggi terhadap pola temporal curah hujan, termasuk lonjakan ekstrem yang penting untuk peringatan dini.

Di sisi lain, meskipun SSA-LSSVR menghasilkan prediksi yang lebih “stabil” dan tidak fluktuatif, model ini cenderung gagal menangkap pola ekstrem. Hal ini bisa jadi karena sifat regresi linier LSSVR yang tidak sebaik LSTM dalam memahami dependensi jangka panjang. Namun demikian, SSA-LSSVR masih berguna untuk aplikasi yang memerlukan kestabilan prediksi, seperti estimasi jangka panjang rata-rata curah hujan.

### 4.2. Manfaat Strategis SSA dan NMF

Kombinasi SSA dengan Non-Negative Matrix Factorization (NMF) dalam preprocessing terbukti meningkatkan keakuratan model secara signifikan. Ini sejalan dengan studi Olushina Olawale Awe [6], yang menekankan pentingnya pendekatan non-negatif untuk data seperti curah hujan. SSA-NMF metode yang efektif dalam mencegah munculnya nilai negatif dalam rekonstruksi deret waktu, menjadikannya lebih sesuai untuk data seperti curah hujan yang secara alami tidak negatif, mampu membersihkan noise dan mempertahankan pola musiman yang penting sebagai input model prediktif.

Rekonstruksi data dengan pendekatan ini memberikan fondasi kuat bagi model LSSVR dan LSTM untuk belajar dari data bersih dan terstruktur, tanpa kehilangan esensi temporal yang penting. Dengan kata lain, kualitas input yang baik akan menentukan kualitas prediksi, dan pendekatan ini menjawab kebutuhan tersebut dengan sangat baik.

### 4.3. Perbandingan dengan Penelitian Sebelumnya

Penelitian ini sejalan dengan hasil dari Hao Liu [35] yang menunjukkan bahwa kombinasi SSA-LSTM merupakan pendekatan paling akurat dalam prediksi deret waktu hidrologi dibandingkan metode hybrid lainnya. Namun yang membedakan studi ini adalah penggunaan data harian dari pos hujan kerjasama di Kota Semarang yang memiliki karakteristik unik seperti pengaruh topografi, iklim pesisir, dan urbanisasi.

Penelitian Arslan Habib [10] dan Jian Zhao [14] juga memperlihatkan keunggulan metode hybrid dalam aplikasi energi dan hidrologi, tetapi pendekatan penelitian ini memperkuatnya dengan visualisasi spasial berbasis GIS, sehingga hasil tidak hanya berhenti pada angka, tetapi dapat dipahami secara geografis. Pendekatan ini menjadi sangat relevan untuk pemangku kebijakan lokal dan lembaga seperti BMKG, BPBD, PUSDATARU dalam merancang intervensi berbasis wilayah.

#### 4.4. Kebutuhan Komputasi (Computational Cost)

Kebutuhan komputasi sangat penting dalam melakukan pengolahan, analisa dan prakiraan curah hujan harian menggunakan model hybrid SSA-LSSVR dan SSA-LSTM agar dapat memaksimalkan hasil sesuai dengan perangkat keras dan lunak serta kecepatan keluaran. Berikut ini adalah beberapa detail dari pengaturan perangkat keras, dataset, kompleksitas algoritma dan informasi waktu pelatihan dari penelitian yang dilakukan :

Tabel 11. Spesifikasi komputasi

| Spesifikasi Komputasi |  |
|-----------------------|--|
| CPU                   | AMD64 Family 23 Model 96 Stepping 1, Authentic AMD RYZEN 5 PRO 4650U with Radeon Graphic |
| Physical Cores        | 6  |
| Logical Cores         | 12   |
| Max Freq Mhz          | 2100 MHz   |
| RAM                   | 16 GB  |
| OS                    | Windows 11 Pro 64-bit  |
| Dataset               | 15 tahun data harian ( $\approx 5.475$ sampel)   |
| Training : 80 %       | ( $\approx 4.380$ sampel)  |
| Testing : 20 %        | ( $\approx 1.095$ sampel)  |

Tabel 12. Kompleksitas komputasi atau kompleksitas algoritma

| Model | Aspek     | Kompleksitas   | Waktu Patokan (s) | Peak Memori Puncak (MiB) |
|-------|-----------|--|-------------------|--------------------------|
| LSSVR | Training  | $O(n^3)$   | 22.17             | 480.7                    |
| LSSVR | Inference | $O(n_{sv} \times d)$                                     | 1.91              | 477.8                    |
| LSTM  | Training  | $O(\text{epochs} \times t\_steps \times \text{units}^2)$ | 7.29              | 679.6                    |
| LSTM  | Inference | $O(t\_steps \times \text{units})$                        | 2.63              | 680.4                    |

Tabel 13. Informasi Pelatihan SSA-LSSVR dan SSA-LSTM

| Model     | Optimasi Parameter | Waktu Tuning (s) | Waktu Training (s) | Parameter Terbaik   |
|-----------|--------------------|------------------|--------------------|---|
| SSA-LSSVR | RandomizedSearchCV | 10.36            | 10.36              | $C=3.315$ , $\text{gamma} = 6.703$ ,<br>kernel = RBF  |
| SSA-LSTM  | Keras Tuner        | 9.11             | 74.69              | Units = [96,64],<br>activation = relu, learning<br>rate = 0.001, batch size =<br>32, epochs = 100 |

Tabel 11 - 13 menyajikan spesifikasi perangkat keras, kompleksitas algoritma, serta waktu dan memori yang digunakan selama pelatihan dan pengujian model hybrid SSA-LSSVR dan SSA-LSTM. Proses eksperimen dilakukan pada perangkat dengan prosesor AMD Ryzen 5 PRO 4650U (6 core fisik, 12 thread), frekuensi 2.1 GHz, RAM 16 GB, dan sistem operasi Windows 11 Pro 64-bit. Dataset terdiri dari 5.475 sampel data hujan harian selama 15 tahun, yang dibagi menjadi 80% data latih dan 20% data uji.

Berdasarkan Tabel 12, secara teoretis LSSVR memiliki kompleksitas pelatihan sebesar  $O(n^3)$  yang menandakan beban komputasi tinggi untuk jumlah data besar akibat inversi matriks kernel. Sebaliknya, LSTM memiliki kompleksitas pelatihan  $O(\text{epochs} \times t\_steps \times \text{units}^2)$  yang bergantung pada jumlah epoch dan unit dalam jaringan, sehingga lebih fleksibel terhadap skala data.

Secara empiris, hasil *benchmark* menunjukkan bahwa waktu pelatihan LSSVR (22,17 detik) lebih lama dibandingkan LSTM (7,29 detik), namun penggunaan memori LSTM (680 MiB) lebih tinggi dibandingkan LSSVR (480 MiB). Hal ini sejalan dengan sifat LSTM yang menyimpan state dan gradient antar langkah waktu.

Tabel 13 menunjukkan proses tuning LSSVR menggunakan *RandomizedSearchCV* membutuhkan waktu 10,36 detik dengan parameter terbaik  $C = 3.315$ ,  $\text{gamma} = 6.703$ , dan kernel RBF. Sementara itu, tuning LSTM menggunakan *Keras Tuner* dengan 9,11 detik menghasilkan konfigurasi optimal: unit LSTM [96, 64], aktivasi ReLU, learning rate 0.001, batch size 32, dan 100 epoch pelatihan.

Secara keseluruhan, LSTM lebih efisien dalam waktu pelatihan, namun memerlukan memori lebih besar dibandingkan LSSVR. Pemilihan model dapat disesuaikan dengan ketersediaan sumber daya komputasi dan kebutuhan implementasi praktis di lapangan.

#### 4.5. Implikasi Praktis dan Akademis

Hasil penelitian ini bukan hanya sekadar pengembangan algoritma, tetapi menjadi bagian dari solusi adaptif terhadap perubahan iklim yang semakin nyata. Model yang dikembangkan dapat diterapkan langsung pada sistem peringatan dini banjir atau kekeringan di tingkat lokal, khususnya di Kota Semarang, dan dengan penyesuaian dapat diperluas ke wilayah lain.

Secara akademis, pendekatan hybrid SSA-NMF dengan LSSVR dan LSTM memperkaya khazanah literatur dalam pemodelan prediksi curah hujan. Studi ini juga menutup celah yang sebelumnya masih kurang dibahas, yaitu integrasi pra-pemrosesan deret waktu dengan deep learning berbasis spasial.

Pemanfaatan kecerdasan buatan untuk prediksi cuaca bukan hanya tentang teknologi, tetapi juga tentang menyelamatkan kehidupan dan masa depan. Harapannya, hasil penelitian ini dapat menjadi kontribusi kecil dalam mengantisipasi bencana hidrometeorologi di Indonesia, dan dapat terus dikembangkan di masa mendatang. Pendekatan model hybrid yang diusulkan dapat diintegrasikan ke dalam platform pengambilan keputusan lokal untuk meningkatkan pemetaan risiko banjir dan alokasi sumber daya air.

## 5. KESIMPULAN

Penelitian ini telah berhasil mengembangkan dan mengevaluasi dua model hybrid, yaitu SSA-LSSVR dan SSA-LSTM, dalam memprediksi curah hujan harian di Kota Semarang. Dengan mengintegrasikan teknik Singular Spectrum Analysis (SSA) yang dipadukan dengan Non-Negative Matrix Factorization (NMF) untuk pra-pemrosesan data, model dapat mengeliminasi noise dan menangkap pola utama dari data curah hujan deret waktu secara efektif.

Berdasarkan hasil evaluasi dengan metrik statistik seperti MSE, RMSE, MAE, dan NSE, diketahui bahwa model SSA-LSTM menunjukkan performa terbaik dibanding SSA-LSSVR di hampir semua pos hujan, terutama dalam menangkap lonjakan ekstrem curah hujan dan pola temporal yang kompleks. Kemampuan LSTM dalam memahami ketergantungan data jangka panjang sangat efektif ketika didukung oleh input data yang telah dibersihkan dan direkonstruksi oleh SSA-NMF.

Model SSA-LSSVR tetap relevan, terutama untuk data dengan pola yang lebih linier dan dalam situasi di mana sumber daya komputasi terbatas. Namun demikian, dari segi keakuratan prediksi dan kemampuan generalisasi terhadap data baru, SSA-LSTM unggul secara konsisten.

Hasil prediksi curah hujan divisualisasikan secara spasial menggunakan teknik interpolasi IDW dan pemetaan dengan GIS, yang memberikan gambaran distribusi hujan yang informatif di wilayah Kota Semarang. Secara umum, studi ini menunjukkan bahwa metode hibrid yang menggabungkan pembelajaran mesin dan pembelajaran dalam dengan langkah awal pra-pemrosesan data waktu dapat meningkatkan akurasi proyeksi hujan dengan drastis. Di samping itu, temuan dari penelitian ini memiliki potensi untuk diterapkan pada daerah lain di Jawa Tengah yang sangat krusial untuk upaya pengurangan risiko bencana hidrometeorologi di daerah tropis.

Secara reflektif, model SSA-NMF-LSTM lebih baik daripada ARIMA karena kemampuan model untuk memahami pola-pola tidak terduga dan tidak teratur seperti fluktuasi tak periodik atau anomali dalam data (*Non-linearitas adaptif*). Sementara ARIMA terbatas pada pola linear, dekomposisi SSA-NMF membagi kompleksitas data menjadi bagian yang dapat diinterpretasikan, yang kemudian dipelajari secara dinamis oleh LSTM. Penanganan ketergantungan temporal jangka panjang dan ekstraksi fitur terarah melalui SSA-NMF, yang mengurangi risiko overfitting pada noise, memberikan keunggulan utama dibandingkan dengan MLP. Sinergi ketiga komponen ini memberikan kerangka kerja yang luas untuk data deret waktu multi-skala, di mana metode statistik konvensional sering gagal karena pendekatan yang terfragmentasi karena pemisahan antara tahap pre-processing dan pemodelan yang tidak saling beradaptasi. Hal ini menyebabkan hilangnya informasi multi-layer pada data kompleks, integrasi metode hybrid SSA-NMF-LSTM adalah solusi atas keterbatasan pendekatan tunggal.

Penelitian lebih lanjut dapat mengeksplorasi skalabilitasnya terhadap kumpulan data regional yang lebih besar atau kemampuan adaptasinya dalam jaringan sensor secara real time. Penelitian di masa depan dapat mengeksplorasi model hibrida berbasis ensemble atau menguji pendekatan ini di wilayah dengan kondisi iklim yang berbeda.

## KONFLIK KEPENTINGAN

Penulis mengungkapkan bahwa tidak terdapat benturan kepentingan yang berkaitan dengan aspek kepengarangan, subjek penelitian, atau penerbitan artikel ini. Setiap fase penelitian dilaksanakan dengan cara mandiri dan tanpa bias, bebas dari pengaruh keuangan, profesional, atau pribadi.

## UCAPAN TERIMA KASIH

Ucapan terimakasih saya sampaikan kepada Stasiun Klimatologi Jawa Tengah untuk data curah hujannya. Pak Eri Zuliarso selaku dosen pembimbing dengan masukan, saran dan bantuannya sehingga terselesaikannya dalam penelitian ini.

## DAFTAR ISI

- [1] M. Alfandi and A. M. H. Sihite, "Penerapan Metode CNN-LSTM Dalam Memprediksi Hujan Pada Wilayah Medan," *Nas. Teknol. Inf. dan Komputer*, vol. 6, no. 1, pp. 490–499, 2022, doi: 10.30865/komik.v6i1.5713.
- [2] S. M. Shahrudin, N. Ahmad, and N. H. Zainuddin, "Modified Singular Spectrum Analysis In Identifying Rainfall Trend Over Peninsular Malaysia," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 15, no. 1, pp. 283–293, 2019, doi: 10.11591/ijeecs.v15.i1.pp283-293.
- [3] P. C. S. Reddy, Y. Sucharitha, and G. S. Narayana, "Development of Rainfall Forecasting Model Using Machine Learning With Singular Spectrum Analysis," *IJUM Eng. J.*, vol. 23, no. 1, pp. 172–186, 2022, doi: 10.31436/IJUM.EJ.V23I1.1822.
- [4] F. Fitri, R. Rahmat, and A. D. Pengestuti, "Forecasting of Rainfall in Sumatera Barat: Singular Spectrum Analysis (SSA) Application," *J. Phys. Conf. Ser.*, vol. 1554, no. 1, 2020, doi: 10.1088/1742-6596/1554/1/012047.
- [5] P. O. Bojang, T. C. Yang, Q. B. Pham, and P. S. Yu, "Linking Singular Spectrum Analysis And Machine Learning For Monthly Rainfall Forecasting," *Appl. Sci.*, vol. 10, no. 9, 2020, doi: 10.3390/app10093224.

- 
- [6] O. O. Awe, R. Mahmoudvand, and P. C. Rodrigues, "Non-Negative Time Series Reconstruction via Singular Spectrum Analysis: A Case Study of Precipitation Dynamics in Nigeria," *Fluct. Noise Lett.*, vol. 19, no. 4, pp. 1–21, 2020, doi: 10.1142/S0219477520500455.
- [7] A. D. Nguyen *et al.*, "Accurate Discharge And Water Level Forecasting Using Ensemble Learning With Genetic Algorithm And Singular Spectrum Analysis-Based Denoising," *Sci. Rep.*, vol. 12, no. 1, pp. 1–25, 2022, doi: 10.1038/s41598-022-22057-8.
- [8] X. Zhang, J. Wang, and Y. Gao, "A Hybrid Short-Term Electricity Price Forecasting Framework: Cuckoo Search-Based Feature Selection With Singular Spectrum Analysis And SVM," *Energy Econ.*, vol. 81, pp. 899–913, 2019, doi: 10.1016/j.eneco.2019.05.026.
- [9] W. Xin, L. Xingke, W. Jiantao, S. Yun, L. Ying, and W. Tingshu, "A Trend Prediction Method for Failures Time Series Data by Exploring Singular Spectrum Analysis and Support Vector Machines Regression," *Proc. IEEE 7th Int. Conf. Comput. Sci. Netw. Technol. ICCSNT 2019*, pp. 535–541, 2019, doi: 10.1109/ICCSNT47585.2019.8962511.
- [10] A. Habib, R. Abbassi, A. J. Aristizábal, and A. Abbassi, "Forecasting Model For Wind Power Integrating Least Squares Support Vector Machine, Singular Spectrum Analysis, Deep Belief Network, And Locality-Sensitive Hashing," *Wind Energy*, vol. 23, no. 2, pp. 235–257, 2020, doi: 10.1002/we.2425.
- [11] H. Zhang *et al.*, "A Combined Model Based On SSA, Neural Networks, And LSSVM For Short-Term Electric Load And Price Forecasting," *Neural Comput. Appl.*, vol. 33, no. 2, pp. 773–788, 2021, doi: 10.1007/s00521-020-05113-0.
- [12] S. Zhang, C. Liu, W. Wang, and B. Chang, "Twin Least Square Support Vector Regression Model Based On Gauss-Laplace Mixed Noise Feature With Its Application In Wind Speed Prediction," *Entropy*, vol. 22, no. 10, pp. 1–19, 2020, doi: 10.3390/e22101102.
- [13] N. Neeraj, J. Mathew, M. Agarwal, and R. K. Behera, "Long Short-Term Memory-Singular Spectrum Analysis-Based Model For Electric Load Forecasting," *Electr. Eng.*, vol. 103, no. 2, pp. 1067–1082, 2021, doi: 10.1007/s00202-020-01135-y.
- [14] J. Zhao, R. Cai, and W. Sun, "Regional Sea Level Changes Prediction Integrated With Singular Spectrum Analysis And Long-Short-Term Memory Network," *Adv. Sp. Res.*, vol. 68, no. 11, pp. 4534–4543, 2021, doi: 10.1016/j.asr.2021.08.017.
- [15] A. Y. Barrera-Animas, L. O. Oyedele, M. Bilal, T. D. Akinosho, J. M. D. Delgado, and L. A. Akanbi, "Rainfall Prediction: A Comparative Analysis Of Modern Machine Learning Algorithms For Time-Series Forecasting," *Mach. Learn. with Appl.*, vol. 7, no. November 2021, p. 100204, 2022, doi: 10.1016/j.mlwa.2021.100204.
- [16] O. A. Wani *et al.*, "Predicting Rainfall Using Machine Learning, Deep Learning, And Time Series Models Across An Altitudinal Gradient In The North-Western Himalayas," *Sci. Rep.*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-77687-x.
- [17] J. Lei, Q. Quan, P. Li, and D. Yan, "Research On Monthly Precipitation Prediction Based On The Least Square Support Vector Machine With Multi-Factor Integration," *Atmosphere (Basel)*, vol. 12, no. 8, 2021, doi: 10.3390/atmos12081076.
- [18] P. Saeipourdizaj, P. Sarbakhsh, and A. Gholampour, "Application Of Imputation Methods For Missing Values Of Pm10 And O3 Data: Interpolation, Moving Average And K-Nearest Neighbor Methods," *Environ. Heal. Eng. Manag.*, vol. 8, no. 3, pp. 215–226, 2021, doi: 10.34172/EHEM.2021.25.
- [19] A. Widiyanti and I. Pratama, "Penanganan Missing Values Dan Prediksi Data Timbunan Sampah Berbasis Machine Learning," vol. 9, no. 2, pp. 242–251, 2024, doi: <https://doi.org/10.36341/rabit.v9i2.4789>.
- [20] A. H. Mauladha and W. Astuti, "Perbandingan Teknik Penanganan Missing Value Dalam Klasifikasi Penyakit Diabetes Menggunakan Metode Naïve Bayes," vol. 2, no. 5, pp. 1–14, 2021.
- [21] Gde Agung Brahmana Suryanegara, Adiwijaya, and Mahendra Dwifabri Purbolaksono, "Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi Pasien Penderita Diabetes Menggunakan Metode Normalisasi," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 114–122, 2021, doi: 10.29207/resti.v5i1.2880.
- [22] S. Smyl, "A Hybrid Method Of Exponential Smoothing And Recurrent Neural Networks For Time Series Forecasting," *Int. J. Forecast.*, vol. 36, no. 1, pp. 75–85, 2020, doi:
-

- 10.1016/j.ijforecast.2019.03.017.
- [23] N. Golyandina, V. Nekrutkin, and A. A. Zhigljavsky, *Analysis of Time Series Structure SSA and Related Techniques*. Boca Raton, Florida: Chapman & Hall/CRC, 2001. doi: 10.1198/004017002320256477.
- [24] Y. Wang, S. Sun, and Z. Cai, "Daily Peak-Valley Electric-Load Forecasting Based on an SSA-LSTM-RF Algorithm," *Energies*, vol. 16, no. 24, 2023, doi: 10.3390/en16247964.
- [25] J. A. K. Suykens and J. Vandewalle, "Least Squares Support Vector Machine Classifiers," *Neural Process. Lett.*, vol. 9, pp. 293–300, 1999, doi: <https://doi.org/10.1023/A:1018628609742>.
- [26] R. Chen, C. Y. Liang, W. C. Hong, and D. X. Gu, "Forecasting Holiday Daily Tourist Flow Based On Seasonal Support Vector Regression With Adaptive Genetic Algorithm," *Appl. Soft Comput.*, vol. 26, pp. 435–443, 2015, doi: 10.1016/j.asoc.2014.10.022.
- [27] J. Huang, Y. Bo, and H. Wang, "Electromechanical Equipment State Forecasting Based On Genetic Algorithm - Support Vector Regression," *Expert Syst. Appl.*, vol. 38, no. 7, pp. 8399–8402, 2011, doi: 10.1016/j.eswa.2011.01.033.
- [28] J. H. Min and Y. C. Lee, "Bankruptcy Prediction Using Support Vector Machine With Optimal Choice Of Kernel Function Parameters," *Expert Syst. Appl.*, vol. 28, no. 4, pp. 603–614, 2005, doi: 10.1016/j.eswa.2004.12.008.
- [29] H. Xiang, Y. Li, H. Liao, and C. Li, "An Adaptive Surrogate Model Based On Support Vector Regression And Its Application To The Optimization Of Railway Wind Barriers," *Struct. Multidiscip. Optim.*, vol. 55, no. 2, pp. 701–713, 2017, doi: 10.1007/s00158-016-1528-9.
- [30] C. Yan, X. Shen, and F. Guo, "An Improved Support Vector Regression Using Least Squares Method," *Struct. Multidiscip. Optim.*, vol. 57, no. 6, pp. 2431–2445, 2018, doi: 10.1007/s00158-017-1871-5.
- [31] H. Nuha, "Mean Squared Error (MSE) dan Penggunaannya," *Papers.Ssrn.Com*, vol. 52, pp. 1–1, 2023, [Online]. Available: <https://ssrn.com/abstract=4420880> Accessed: Jun. 18, 2025
- [32] T. O. Hodson, "Root-Mean-Square Error (RMSE) Or Mean Absolute Error (MAE): When To Use Them Or Not," *Geosci. Model Dev.*, vol. 15, no. 14, pp. 5481–5487, 2022, doi: 10.5194/gmd-15-5481-2022.
- [33] H. V. Gupta and H. Kling, "On Typical Range, Sensitivity, And Normalization Of Mean Squared Error And Nash-Sutcliffe Efficiency Type Metrics," *Water Resour. Res.*, vol. 47, no. 10, pp. 2–4, 2011, doi: 10.1029/2011WR010962.
- [34] R. le Roux, S. Henrico, J. Bezuidenhout, and I. Henrico, "Inverse Distance Weighting As An Alternative Interpolation Method To Create Radiometric Maps Of Natural Radionuclide Concentrations Using QGIS," *Proc. ICA*, vol. 5, no. August, pp. 1–7, 2023, doi: 10.5194/ica-proc-5-10-2023.
- [35] H. Liu, W. Liu, J. Luo, and J. Li, "Comparative Study of Multi-Combination Models for Medium- and Long-Term Runoff Prediction in Weihe River," *IEEE Access*, vol. 11, pp. 97099–97106, 2023, doi: 10.1109/ACCESS.2023.3312185.

