

Evaluation of ARIMA, ARIMA-LSTM, LSTM, and CNN-LSTM Models for Daily Air Quality Index Forecasting in Jakarta

Mushliha^{1*}, Nisrinah²

¹ Teknologi Informasi, Teknik & Informatika, Universitas Bina Sarana Informatika, DKI Jakarta

² Kesehatan Masyarakat, Ilmu – ilmu Kesehatan, Universitas Esa Unggul, DKI Jakarta

Email: ^{1*} mushliha.ush@bsi.ac.id, ² nisrinah@esaunggul.ac.id

(*: corresponding author)

Abstract - Increasing air pollution in urban areas, particularly in DKI Jakarta, has made a reliable air quality prediction system increasingly essential for environmental control and public health risk management. The Air Quality Index (AQI) exhibits complex and fluctuating patterns, requiring forecasting methods capable of capturing both linear and non-linear. This study aims to conduct a comparative analysis of statistical, deep learning, and hybrid models for AQI forecasting using daily AQI data from Jakarta during the 2023–2025 period. The dataset includes pollutant parameters such as $PM_{2.5}$, PM_{10} , NO_2 , SO_2 , O_3 , and CO. The proposed models consist of Autoregressive Integrated Moving Average (ARIMA), Long Short-Term Memory (LSTM), Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM), and a hybrid ARIMA-LSTM model. The research methodology includes data preprocessing, normalization using Min-Max Scaling, sequence generation using the sliding window approach, model training, and evaluation using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The experimental results show that the CNN-LSTM model achieved the best forecasting performance with MAE, RMSE, and MAPE values of 4.54, 6.03, and 13.83%, respectively, followed closely by the LSTM model. Meanwhile, the ARIMA model produced the lowest performance, and the hybrid ARIMA-LSTM model did not outperform the standalone deep learning models. These findings indicate that deep learning approaches, particularly CNN-LSTM, are more effective in capturing the complex dynamics of urban air pollution data and have strong potential to support air quality forecasting and pollution control systems in Jakarta.

Keywords: Air Quality Index (AQI), ARIMA, Air Quality Forecasting, CNN-LSTM, Deep Learning.

Abstrak – Peningkatan pencemaran udara di wilayah perkotaan, khususnya DKI Jakarta, menyebabkan kebutuhan akan sistem prediksi kualitas udara yang andal semakin penting dalam pelaksanaan pengendalian lingkungan dan manajemen risiko kesehatan masyarakat. Indeks Standar Pencemar Udara (ISPU) memiliki karakteristik data yang fluktuatif dan kompleks sehingga memerlukan metode prediksi yang mampu menangkap pola linear maupun non-linear. Penelitian ini bertujuan untuk melakukan analisis komparatif terhadap model statistik, *deep learning*, dan *hybrid* dalam peramalan ISPU menggunakan data harian ISPU DKI Jakarta tahun 2023–2025 dengan parameter polutan $PM_{2.5}$, PM_{10} , NO_2 , SO_2 , O_3 , dan CO. Model yang digunakan meliputi *Autoregressive Integrated Moving Average* (ARIMA), *Long Short-Term Memory* (LSTM), *Convolutional Neural Network–Long Short-Term Memory* (CNN-LSTM), dan ARIMA-LSTM. Tahap penelitian meliputi *preprocessing* data, normalisasi menggunakan *Min-Max Scaling*, pembentukan data deret waktu berbasis jendela pengamatan (*sliding window*), pelatihan model, serta evaluasi menggunakan *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE). Hasil penelitian menunjukkan bahwa model CNN-LSTM memberikan performa terbaik dengan nilai MAE sebesar 4.54, RMSE sebesar 6.03, dan MAPE sebesar 13.83%, diikuti oleh model LSTM dengan performa yang kompetitif. Sementara itu, model ARIMA menghasilkan performa terendah dan model ARIMA-LSTM belum mampu melampaui performa model *deep learning* tunggal. Hasil penelitian menunjukkan bahwa model *deep learning*, khususnya CNN-LSTM, lebih efektif dalam menangkap pola kompleks kualitas udara perkotaan dan berpotensi mendukung sistem prediksi serta pengendalian pencemaran udara di DKI Jakarta.

Kata Kunci: ARIMA, CNN-LSTM, Deep Learning, Indeks Kualitas Udara, Peramalan Kualitas Udara.

1. PENDAHULUAN

DKI Jakarta menghadapi permasalahan pencemaran udara yang termasuk kategori kritis di Indonesia. Statusnya sebagai ibu kota dengan kepadatan penduduk dan volume lalu lintas yang tinggi berkontribusi terhadap penurunan kualitas udara. Selain itu, Jakarta sering mencatatkan Indeks Standar Pencemaran Udara (ISPU) pada kategori tidak sehat hingga sangat tidak sehat, terutama pada musim kemarau [1]. Paparan polutan seperti $PM_{2.5}$, PM_{10} , NO_2 , SO_2 , O_3 , dan CO yang tinggi dapat menyebabkan peningkatan kasus penyakit pernapasan, gangguan kardiovaskular, serta kerugian ekonomi yang signifikan [2]. Berbagai kebijakan telah diterapkan untuk mengurangi pencemaran udara namun pencemaran udara di DKI Jakarta masih cukup sulit dikendalikan sepenuhnya sehingga indeks kualitas udara (IKU) belum mencapai kategori yang baik [3]. Indeks Standar Pencemar Udara (ISPU) digunakan sebagai indikator utama dalam memantau kondisi kualitas udara secara harian. Namun, dinamika ISPU yang bersifat fluktuatif dan dipengaruhi oleh berbagai faktor kompleks sehingga menyebabkan kesulitan dalam melakukan prediksi yang presisi sehingga dibutuhkan suatu pemodelan yang dapat menangkap pola data yang kompleks agar dapat mendukung sistem pengendalian pencemaran udara secara lebih efektif.

Beberapa tahun terakhir teknik *machine learning* dan *deep learning* digunakan untuk data temporal yang kompleks [4]. Kedua teknik ini juga banyak diterapkan untuk memprediksi kualitas udara dan menunjukkan performa yang baik [5]. Metode statistik konvensional seperti *Autoregressive Integrated Moving Average* (ARIMA) telah lama

digunakan karena mampu mendeteksi pola linear dan tren dalam data deret waktu [6]. Namun, performa ARIMA dapat menurun ketika data menunjukkan pola yang kompleks dan non-linear, sehingga diperlukan pendekatan lain yang mampu menangkap karakteristik tersebut secara lebih efektif [7]. *Model deep learning* seperti *Long Short-Term Memory* (LSTM) semakin banyak digunakan karena kemampuannya dalam memahami ketergantungan jangka panjang dan pola non-linear [8]. Hasil penelitian menunjukkan bahwa model LSTM mampu memberikan hasil yang lebih baik dibandingkan ARIMA dalam peramalan kualitas udara di wilayah perkotaan [8]. Perkembangan penelitian selanjutnya menggabungkan ARIMA-LSTM untuk meningkatkan performa model prediksi untuk $PM_{2.5}$ hasil penelitian, menunjukkan peningkatan model yang signifikan [9]. Meskipun efektif, model ARIMA-LSTM memerlukan penyesuaian parameter model yang tepat untuk mencapai kinerja prediksi yang optimal [10].

Pada penelitian yang mengembangkan model *Convolutional Neural Network*-LSTM (CNN-LSTM) yang menggabungkan kemampuan ekstraksi fitur lokal dari CNN dan pemodelan temporal dari LSTM menghasilkan peningkatan akurasi prediksi. Penggunaan model CNN-LSTM memiliki tantangan seperti *overfitting* dan penyetelan parameter namun dapat diatasi dengan menggunakan teknik optimisasi. Berdasarkan penelitian terdahulu, pendekatan kombinasi seperti ARIMA-LSTM dan CNN-LSTM dapat meningkatkan performa prediksi dan mampu meningkatkan akurasi pada data *time series* yang memiliki kombinasi pola linear dan non-linear [11].

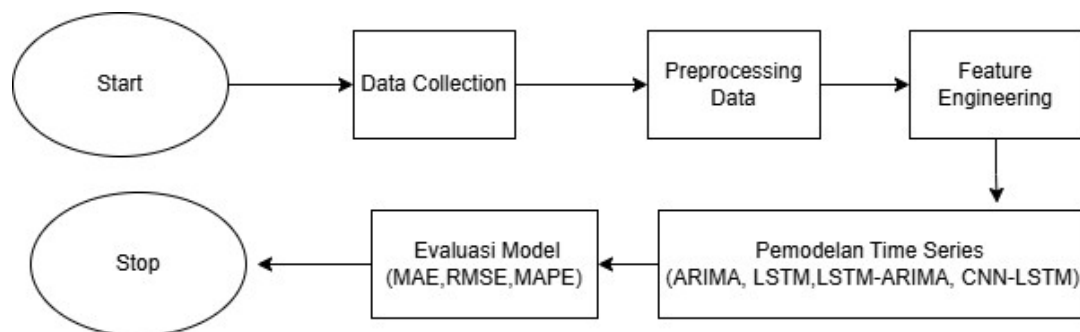
Banyak riset menunjukkan keunggulan penggunaan model *deep learning* dan kombinasinya untuk meramal kualitas udara, namun penerapan model-model tersebut masih belum sempurna. Pertama, sebagian besar penelitian dilakukan pada dataset yang mencakup variabel meteorologi tambahan sehingga belum banyak kajian yang secara khusus mengevaluasi performa model hanya berdasarkan data ISPU. Kedua, belum banyak penelitian yang secara komprehensif membandingkan model *deep learning* (seperti LSTM dan CNN-LSTM), model statistik (ARIMA), dan model *hybrid* (ARIMA-LSTM) dalam satu kerangka evaluasi yang sama, khususnya pada konteks kota besar di Indonesia seperti Jakarta. Ketiga, efektivitas model kombinasi dalam meningkatkan performa prediksi masih menunjukkan hasil yang bervariasi, sehingga diperlukan kajian lebih lanjut untuk memahami kondisi di mana model *hybrid* benar-benar memberikan peningkatan performa.

Berdasarkan paparan di atas tersebut, penelitian ini bertujuan untuk melakukan evaluasi kemampuan model ARIMA, *deep learning* dan model *hybrid* dalam peramalan ISPU di DKI Jakarta menggunakan data historis ISPU. Model yang digunakan dalam penelitian ini meliputi CNN-LSTM, LSTM, ARIMA, dan model ARIMA-LSTM. Penelitian ini diharapkan dapat memperkaya wawasan komprehensif terkait performa masing-masing model dalam menangkap dinamika kualitas udara perkotaan, serta memberikan rekomendasi model terbaik yang dapat digunakan untuk mendukung pengendalian pencemaran udara secara lebih efektif. Selain itu, Temuan dari penelitian ini dapat dijadikan sebagai landasan dalam merancang sistem peramalan kualitas udara yang memiliki tingkat presisi tinggi serta fungsionalitas yang lebih baik di masa mendatang.

2. METODE PENELITIAN

2.1. Tahap Penelitian

Penelitian ini menerapkan metode kuantitatif dengan memanfaatkan pemodelan deret waktu untuk memproyeksikan angka Indeks Standar Pencemar Udara (ISPU) di wilayah DKI Jakarta. Kerangka penelitian ini dirancang secara sistematis, yang diawali dengan tahap akuisisi data dan diakhiri dengan proses evaluasi model.



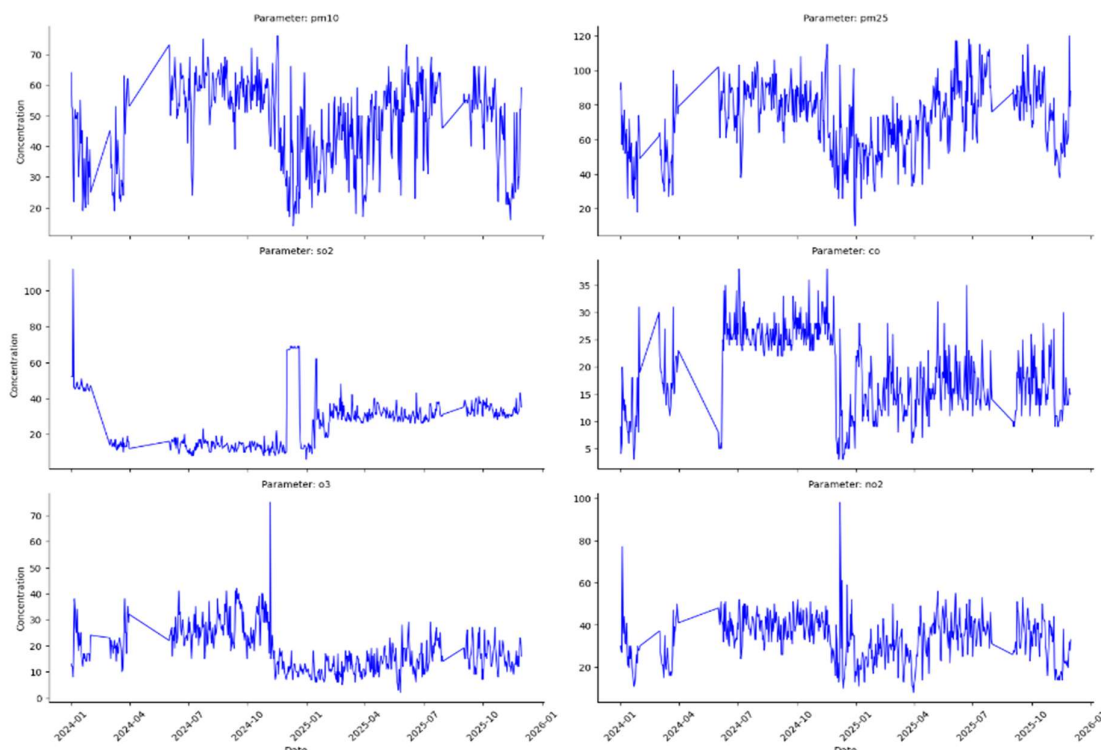
Gambar 1. Alur Tahapan Penelitian

Gambar 1 menyajikan alur tahapan penelitian yang digunakan untuk membangun dan mengevaluasi model peramalan kualitas udara di DKI Jakarta. Tahapan penelitian dirancang secara sistematis mulai dari pengumpulan data hingga evaluasi hasil prediksi untuk memastikan bahwa model yang dihasilkan mampu memberikan performa yang optimal dalam memprediksi konsentrasi polutan

2.2. Data Collection

Adapun data yang digunakan dalam penelitian ini yaitu Data Indeks Standar Pencemar Udara (ISPU) harian dan berasal dari Stasiun Pemantau Kualitas Udara (SPKU) di Provinsi DKI Jakarta mulai 01-01-2023 sampai 30-11-2025 yang diperoleh dari [Satudata Jakarta](#). Dataset ini mencakup beberapa parameter polutan yaitu $PM_{2.5}$, PM_{10} , NO_2 , SO_2 , O_3 , dan CO. Dimana $PM_{2.5}$ adalah partikulat berukuran lebih kecil dari 2.5 mikron, PM_{10} adalah partikulat berukuran lebih kecil dari 10 mikron, NO_2 merupakan parameter nitrogen dioksida, SO_2 adalah parameter sulfur dioksida, CO adalah parameter karbon monoksida, O_3 adalah parameter ozon.

Daily Fluctuation Trends of Pollutant Parameters at DK11 Bundaran Hotel Indonesia (HI)



Gambar 2. Tren Fluktuasi Polutan di Stasiun Bundaran HI

Gambar 2 menyajikan gambaran tren fluktuasi dari semua polutan untuk stasiun Bundaran HI. Data stasiun yang digunakan dalam penelitian ini adalah DK11 Bundaran Hotel Indonesia (HI), DK12 Kelapa Gading, DK13 Jagakarsa, DK14 Lubang Buaya, dan DK15 Kebon Jeruk Jakarta Barat.

2.3. Preprocessing Data

Setelah data memperoleh data dan sebelum memulai pemodelan, dilakukan langkah prapemrosesan untuk menyaring dan menyiapkan data agar memenuhi standar kualitas yang diperlukan. Dalam penelitian ini dilakukan beberapa tahapan *preprocessing data*. Pertama, pembersihan data terdiri dari pengecekan data duplikat dan melakukan standarisasi dari nama stasiun. Gambar 3 menampilkan data yang sudah dilakukan pembersihan missing value, data ganda, dan standarisasi nama stasiun.

Tahap kedua merupakan normalisasi data. Untuk menyeragamkan skala data, penelitian ini menggunakan metode *MinMaxScaler* yang mentransformasikan seluruh nilai ke dalam interval $[0, 1]$. Normalisasi dilakukan dengan rumus persamaan (1):

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Dimana X_{scaled} adalah nilai yang sudah dilakukan normalisasi, X_{min} adalah nilai data paling kecil, X_{max} adalah nilai data paling besar. Penskalaan ini memastikan bahwa semua fitur berkontribusi secara merata terhadap proses pembelajaran model dan membantu mencegah masalah yang disebabkan oleh perbedaan besar dalam besaran fitur [4].

Head of the DataFrame after cleaning:

	date	pm10	pm25	so2	co	o3	no2	categori	stasiun	daily_score
0	2023-01-01	44.0	55.0	47.0	10.0	24.0	9.0	SEDANG	DKI1 Bundaran Hotel Indonesia (HI)	1
1	2023-01-02	32.0	43.0	52.0	9.0	24.0	8.0	SEDANG	DKI1 Bundaran Hotel Indonesia (HI)	1
2	2023-01-03	31.0	35.0	49.0	9.0	12.0	7.0	BAIK	DKI1 Bundaran Hotel Indonesia (HI)	2
3	2023-01-04	30.0	47.0	53.0	11.0	15.0	9.0	SEDANG	DKI1 Bundaran Hotel Indonesia (HI)	1
4	2023-01-05	38.0	50.0	50.0	13.0	26.0	11.0	BAIK	DKI1 Bundaran Hotel Indonesia (HI)	2

Tail of the DataFrame after cleaning:

	date	pm10	pm25	so2	co	o3	no2	categori	stasiun	daily_score
5165	2025-11-26	23.0	58.0	28.0	5.0	19.0	18.0	SEDANG	DKI5 Kebon Jeruk Jakarta Barat	1
5166	2025-11-27	31.0	64.0	31.0	6.0	29.0	21.0	SEDANG	DKI5 Kebon Jeruk Jakarta Barat	1
5167	2025-11-28	48.0	72.0	37.0	9.0	37.0	37.0	SEDANG	DKI5 Kebon Jeruk Jakarta Barat	1
5168	2025-11-29	38.0	66.0	33.0	6.0	30.0	24.0	SEDANG	DKI5 Kebon Jeruk Jakarta Barat	1
5169	2025-11-30	51.0	84.0	31.0	9.0	34.0	31.0	SEDANG	DKI5 Kebon Jeruk Jakarta Barat	1

Gambar 3. Data Bersih

Tahap ketiga adalah pembagian data, penelitian ini menggunakan *TimeSeriesSplit* untuk pembagian data. Pembagian data untuk pelatihan dan evaluasi model menggunakan *TimeSeriesSplit* dari *sklearn.model_selection* karena sangat sesuai untuk data *time series* serta mempertahankan urutan temporal data. Ini bekerja dengan membagi data menjadi beberapa *fold* (lipatan). Pada setiap *fold* data pelatihan terdiri dari bagian awal dari deret waktu. Data pengujian terdiri dari bagian deret waktu yang langsung mengikuti data pelatihan. Strategi ini memungkinkan model berlatih pada rekaman historis dan diuji pada data berikutnya secara berurutan, guna mensimulasikan kondisi prediksi dunia nyata. Hal ini akan memberikan evaluasi yang lebih realistis untuk peramalan deret waktu dibandingkan dengan pembagian acak biasa [12]. Penelitian ini menggunakan *n_splits=5*, yang berarti ada 5 set pelatihan dan pengujian yang berbeda, masing-masing dengan data pengujian yang lebih baru dari set pelatihan.

2.4. Feature Engineering

Tahap *feature engineering* dilakukan menggunakan *sliding window*. Pendekatan *sliding window* berfungsi sebagai mekanisme pembentukan data urutan dari data deret waktu asal, sehingga dapat diproses secara optimal oleh model *deep learning*, khususnya LSTM dan CNN-LSTM. Metode ini digunakan untuk menangkap pola temporal dari data ISPU berdasarkan beberapa waktu sebelumnya (*lookback window*). Gambar 4 menampilkan data yang sudah dilakukan *feature engineering* dan siap untuk digunakan untuk membangun model.

```
First X_train_final:
array([[0.30952381],
       [0.19940476],
       [0.0952381 ]])

First y_train_final:
array([0.23511905])
First X_test_final:
array([[0.38095238],
       [0.66666667],
       [0.61904762]])

First y_test_final:
array([0.65178571])
```

Gambar 4. Data Setelah Feature Engineering

Metode *sliding window* bekerja dengan cara mengambil sejumlah data historis sebelumnya sebagai *input* untuk memprediksi nilai pada waktu berikutnya. Dalam penelitian ini digunakan pemilihan *window size* melalui *hyperparameter tuning* dengan pilihan 3, 5, dan 7. *Window size* 3 berarti model menggunakan data tiga hari sebelumnya untuk memprediksi nilai ISPU pada hari berikutnya.

Secara matematis, proses *sliding window* dapat dinyatakan pada persamaan (2) dan (3):

$$X_t = [x_{t-3}, x_{t-2}, x_{t-1}] \quad (2)$$

$$y_t = x_t \quad (3)$$

dimana X_t merupakan sequence input, $x_{t-3}, x_{t-2}, x_{t-1}$ adalah data historis, y_t merupakan target prediksi pada waktu ke- t [13].

2.5. Pemodelan Time Series

Penelitian ini menggunakan empat model utama, yaitu model statistik, *deep learning*, dan *hybrid*. Model yang dibangun dalam penelitian ini adalah ARIMA, LSTM, ARIMA-LSTM dan CNN-LSTM.

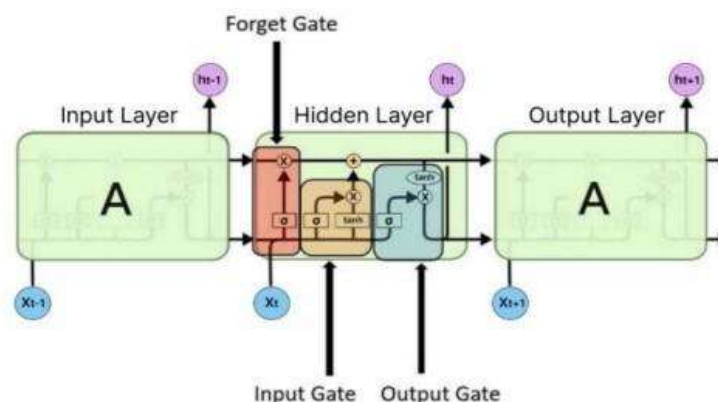
a. Autoregressive Integrated Moving Average (ARIMA)

Autoregressive Integrated Moving Average (ARIMA) adalah metode statistik yang digunakan secara luas untuk analisis deret waktu dan peramalan. Model ini sangat populer karena fleksibilitasnya dalam menangani data deret waktu yang bersifat musiman maupun non-musiman [14]. Model ARIMA digunakan untuk menangkap pola linear dalam data *time series*. Model ini memiliki tiga parameter utama: $ARIMA(p, d, q)$

Dimana ARIMA memiliki tiga komponen utama. Pertama adalah AR (*AutoRegressive*) menggunakan hubungan ketergantungan antara observasi saat ini dengan sejumlah observasi di masa lalu (nilai-nilai *lagged*). Parameter p menunjukkan jumlah observasi masa lalu yang disertakan. Kedua adalah I (*Integrated*) menggunakan operasi diferensiasi untuk membuat deret waktu stasioner. Memiliki rata-rata, varians, dan struktur autokorelasi yang konstan seiring waktu. Parameter d menunjukkan jumlah diferensiasi yang diperlukan. Ketiga adalah MA (*Moving Average*) menggunakan ketergantungan antara observasi saat ini dengan kesalahan prediksi (residu) dari model rata-rata bergerak yang diterapkan pada observasi di masa lalu. Parameter q menunjukkan jumlah kesalahan prediksi masa lalu yang disertakan.

b. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) merupakan salah satu pengembangan dari *Recurrent Neural Network* (RNN) yang dirancang untuk memodelkan data deret waktu dan menangkap ketergantungan jangka panjang dalam suatu rangkaian data [15]. LSTM memiliki tiga komponen utama yaitu *memory cell*, *hidden states*, dan *gate mechanism*. *Memory cell* merupakan komponen inti yang menyimpan informasi jangka panjang. *Memory cell* memungkinkan LSTM untuk mengingat pola dalam data sekuensial. *Hidden States* digunakan untuk membuat prediksi pada langkah waktu tertentu. *Hidden states* bekerja bersama *memory cell* untuk memproses informasi [16]. LSTM memiliki tiga jenis gerbang (*input*, *forget*, dan *output*) yang mengontrol aliran informasi di dalam unit LSTM. Berikut arsitektur LSTM:



Gambar 5. Arsitektur Komponen LSTM [17]

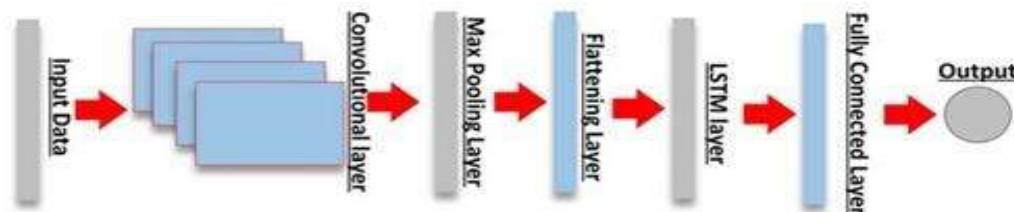
Gambar 5 menyajikan komponen utama pada arsitektur LSTM yang digunakan dalam penelitian ini untuk memodelkan data deret waktu kualitas udara. LSTM mampu mengatasi masalah *vanishing gradient* melalui tiga gerbang (*gates*): *forget gate*, *input gate*, dan *output gate*.

c. ARIMA-LSTM

ARIMA-LSTM adalah model gabungan yang mengintegrasikan pendekatan statistik tradisional ARIMA (*AutoRegressive Integrated Moving Average*) dengan kemampuan pembelajaran mendalam dari LSTM untuk meningkatkan akurasi peramalan data deret waktu [18]. Pada model *hybrid* ini, ARIMA digunakan untuk memproses data historis dan menghasilkan prediksi awal serta residual (sisir error dari prediksi). Residual dari ARIMA dianalisis lebih lanjut menggunakan LSTM untuk menangkap pola non-linear yang tidak dapat ditangkap oleh ARIMA. Hasil dari kedua model digabungkan untuk menghasilkan prediksi akhir yang lebih akurat [19].

d. CNN-LSTM

CNN-LSTM adalah model hibrida yang menggabungkan *Convolutional Neural Networks* (CNN) dan *Long Short-Term Memory* (LSTM) untuk menangani data deret waktu dengan lebih efektif. Model ini memanfaatkan kekuatan masing-masing komponen untuk meningkatkan akurasi dan stabilitas prediksi. CNN digunakan untuk ekstraksi fitur lokal dari data [20], seperti pola spasial atau hubungan antar variabel dalam data deret waktu. Kemudian hasil ekstraksi fitur akan diteruskan ke LSTM. LSTM bertugas untuk menangkap ketergantungan temporal atau pola jangka panjang dalam data deret waktu [21].



Gambar 6. Asitektur CNN-LSTM [16]

Gambar 6 menyajikan arsitektur model CNN-LSTM yang digunakan dalam penelitian untuk melakukan peramalan kualitas udara berdasarkan data historis polutan. Model ini menggabungkan CNN dalam mengekstraksi fitur penting dari data deret waktu dan LSTM dalam mempelajari hubungan temporal jangka pendek maupun jangka panjang. Kombinasi kedua metode tersebut diharapkan mampu meningkatkan akurasi prediksi dibandingkan penggunaan model tunggal. Arsitektur CNN-LSTM yang digunakan dalam penelitian ini menggabungkan lapisan CNN dengan lapisan LSTM untuk memproses data deret waktu.

2.6. Evaluasi Model

Evaluasi model pada penelitian ini menggunakan beberapa metrik untuk membandingkan performa model yang sudah dibangun. Penggunaan beberapa metrik evaluasi dalam penelitian ini bertujuan untuk memperoleh penilaian performa model yang lebih komprehensif dan objektif. Setiap metrik memiliki karakteristik dan fokus evaluasi yang berbeda. *Mean Absolute Error* (MAE) digunakan untuk mengukur rata-rata selisih absolut antara nilai aktual dan hasil prediksi tanpa memperhatikan arah kesalahan. Metrik ini memberikan gambaran umum mengenai besar kesalahan prediksi model secara langsung dalam satuan asli data. Persamaan (4) menyajikan rumus untuk menghitung MAE:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4)$$

Dimana y_i merupakan nilai aktual, $\hat{y}_i$ merupakan nilai prediksi dari model, dan n merupakan jumlah data. Sementara itu, *Root Mean Square Error* (RMSE) digunakan karena lebih sensitif terhadap kesalahan prediksi yang besar (*large error*). Dalam konteks kualitas udara, penggunaan RMSE menjadi penting karena kesalahan prediksi pada kondisi polusi tinggi dapat berdampak signifikan terhadap sistem peringatan dini dan pengambilan keputusan lingkungan. Nilai RMSE yang kecil menunjukkan bahwa model mampu meminimalkan kesalahan ekstrem dalam prediksi ISPU. Persamaan (5) menyajikan rumus untuk menghitung RMSE:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

Di sisi lain, *Mean Absolute Percentage Error* (MAPE) digunakan untuk mengukur tingkat kesalahan dalam bentuk persentase sehingga lebih mudah diinterpretasikan. Metrik ini membantu dalam memahami seberapa besar rata-rata kesalahan prediksi model dibandingkan dengan nilai aktual. Penggunaan MAPE memudahkan perbandingan performa antar model karena dinyatakan dalam satuan persen. Persamaan (6) menyajikan rumus untuk menghitung MAPE:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (6)$$

Dengan mengombinasikan MAE, RMSE, dan MAPE, evaluasi model dapat dilakukan secara lebih menyeluruh, baik dari sisi rata-rata kesalahan, sensitivitas terhadap error besar, maupun tingkat akurasi dalam bentuk persentase. Pendekatan ini memungkinkan penentuan model terbaik dilakukan secara lebih objektif dan dapat memberikan gambaran yang lebih kuat mengenai kemampuan model dalam memprediksi kualitas udara di DKI Jakarta.

3. HASIL DAN PEMBAHASAN

Penelitian ini mengimplementasikan beberapa model *time series forecasting* untuk melakukan prediksi Indeks Standar Pencemar Udara (ISPU) di DKI Jakarta menggunakan data harian tahun 2023–2025. Model yang digunakan meliputi model statistik, *deep learning*, dan *hybrid* yaitu ARIMA, LSTM, ARIMA-LSTM, dan CNN-LSTM. Parameter yang digunakan untuk membangun model ARIMA yaitu Orde (p, d, q) ditentukan secara otomatis oleh *pmdarima.auto_arima* untuk setiap polutan dan setiap lipatan validasi silang. Rentang pencarian untuk p dan q (*start_p=1, max_p=3, start_q=1, max_q=3*). Model ini menggunakan non-musiman (*m=1, seasonal=False*). *Optimizer* yang digunakan dikelola secara internal oleh proses penyesuaian ARIMA. *Loss Function* yang digunakan adalah *Maximum Likelihood Estimation* (MLE). Pada model LSTM parameter yang digunakan *window size* diatur sebanyak 3, units pada lapisan LSTM diatur sebanyak 15, *epochs* diatur 50, *batch_size* diatur 32, fungsi aktivasi (relu) untuk lapisan LSTM dan adam sebagai *optimizer*, serta *mean_squared_error* sebagai fungsi *loss*. Model gabungan ARIMA-LSTM menggunakan semua parameter yang digunakan oleh masing – masing model ARIMA dan LSTM yang sebelumnya berdiri sendiri – sendiri. Selanjutnya untuk model CNN-LSTM, parameter yang digunakan untuk LSTM sama dengan parameter model LSTM yang sebelumnya berdiri sendiri kemudian parameter untuk CNN yang digunakan adalah *filters* (*Conv1D Filters*) diatur 32, *kernel_size* (*Conv1D Kernel Size*) diatur 2, dan *pool size* (*MaxPooling1D Pool Size*) diatur 1.

3.1. Hasil Implementasi Model

Penelitian ini mengimplementasikan model ARIMA, LSTM, ARIMA-LSTM, dan CNN-LSTM untuk peramalan parameter polutan $PM_{2,5}$, PM_{10} , NO_2 , SO_2 , O_3 , dan CO Semua nilai dihitung pada data uji menggunakan *TimeSeriesSplit cross-validation*. Tabel 1 menunjukkan hasil evaluasi model menggunakan metrik MAE, RMSE, dan MAPE. Tabel 1 menyajikan hasil evaluasi performa empat model peramalan, yaitu ARIMA, LSTM, ARIMA-LSTM, dan CNN-LSTM pada enam parameter polutan yang digunakan dalam penelitian, meliputi $PM_{2,5}$, PM_{10} , NO_2 , SO_2 , O_3 , dan CO. Evaluasi dilakukan menggunakan tiga metrik, yaitu *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), dan *Mean Absolute Percentage Error* (MAPE). Semakin kecil nilai ketiga metrik tersebut, semakin baik kemampuan model dalam memprediksi konsentrasi polutan.

Tabel 1. Hasil Evaluasi

Polutan	Model	MAE	RMSE	MAPE (%)
PM_{10}	ARIMA	11.458	14.561	25.549
	LSTM	6.229	8.101	14.543
	ARIMA-LSTM	7.977	9.978	18.218
	CNN-LSTM	6.288	8.131	14.828
$PM_{2,5}$	ARIMA	16.815	21.607	25.475
	LSTM	9.766	12.507	15.676
	ARIMA-LSTM	11.934	15.043	18.660
	CNN-LSTM	9.776	12.584	15.629
SO_2	ARIMA	10.313	15.629	37.569
	LSTM	1.904	3.122	6.847
	ARIMA-LSTM	5.827	8.944	21.523
	CNN-LSTM	1.698	2.704	5.766
CO	ARIMA	3.262	4.282	25.634
	LSTM	2.342	3.268	16.536
	ARIMA-LSTM	2.637	3.539	19.924
	CNN-LSTM	2.397	3.328	16.692
O_3	ARIMA	6.029	7.505	27.065
	LSTM	3.256	4.152	14.610
	ARIMA-LSTM	4.250	5.251	19.238
	CNN-LSTM	3.091	3.992	13.633
NO_2	ARIMA	5.442	6.925	24.760
	LSTM	4.040	5.384	16.514
	ARIMA-LSTM	4.478	5.710	19.434
	CNN-LSTM	3.912	5.234	15.982

Berdasarkan hasil evaluasi yang menyeluruh untuk semua polutan, ARIMA menghasilkan evaluasi paling buruk dengan rata – rata MAE 8.89, RMSE 11.75 dan MAPE 27,68 %. Disusul oleh ARIMA – LSTM dengan MAE 6.23, RMSE 8.10, dan MAPE 19.56. Kemudian LSTM dengan MAPE 4.63, RMSE 6.11, dan, MAPE 14.16. CNN- LSTM merupakan model yang paling unggul dengan MAE 4.49, RMSE 5.96, dan MAPE 13.62%. ARIMA dalam penelitian ini digunakan sebagai *baseline*. Perbandingan performa ketiga model dibandingkan

dengan ARIMA mengalami penurunan untuk ARIMA-LSTM dengan MAE, RMSE, dan MAPE turun sebanyak 29.9%, 31.1%, 29.3%, LSTM dengan MAE, RMSE, dan MAPE turun sebanyak 47.9%, 48.0%, 48.8%, dan CNN-LSTM 49.5%, 49.3%, 50.8%.

Model CNN-LSTM memperoleh rata-rata performa terbaik di seluruh polutan dengan nilai MAE sebesar 4.49, RMSE 5.96, dan MAPE 13.62%. Model ini mengungguli LSTM (peringkat kedua) dan secara signifikan mengalahkan ARIMA-LSTM serta ARIMA. Penurunan error mencapai hampir 50% dibandingkan ARIMA menunjukkan superioritas pendekatan *deep learning* dalam menangani karakteristik data ISPU yang non-stasioner dan non-linear di Jakarta. Hasil ini membuktikan bahwa arsitektur CNN-LSTM lebih sesuai untuk peramalan kualitas udara di Jakarta karena memiliki kemampuan ekstraksi fitur spasial dan pemodelan temporal jangka panjang.

3.2. Analisis Perbandingan Model

Hasil penelitian memperlihatkan bahwa model *deep learning* memiliki kemampuan yang lebih unggul daripada model statistik tradisional dalam peramalan ISPU di DKI Jakarta. Kondisi ini menunjukkan bahwa data kualitas udara memiliki pola yang rumit dan tidak linier sehingga lebih sesuai dimodelkan menggunakan pendekatan *deep learning*. Tabel 2 menyajikan ringkasan model terbaik untuk masing-masing parameter polutan berdasarkan hasil evaluasi menggunakan MAE, RMSE, dan MAPE. Tabel ini bertujuan untuk memberikan gambaran mengenai kemampuan setiap model dalam memprediksi konsentrasi polutan yang berbeda, sehingga dapat diketahui model yang paling sesuai untuk setiap jenis polutan.

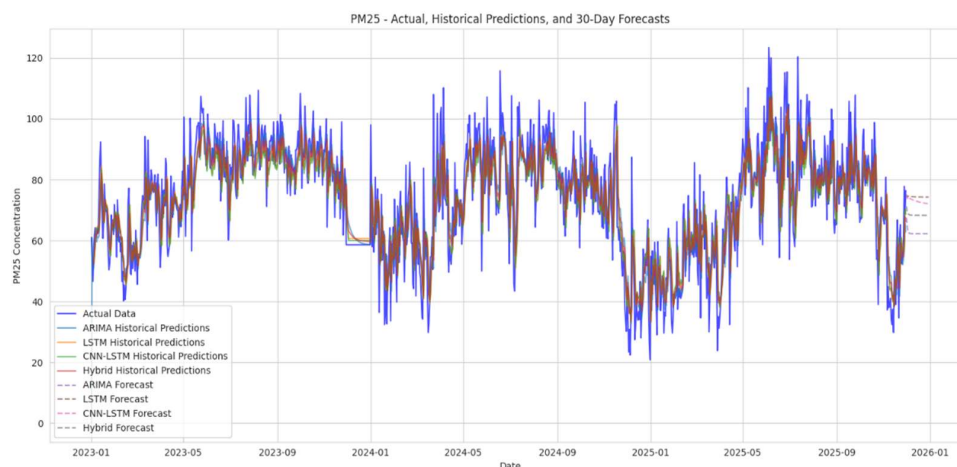
Tabel 2. Analisis Evaluasi Semua Polutan

Polutan	Model Terbaik	MAE	RMSE	MAPE (%)	Keterangan
PM_{10}	LSTM	6.23	8.10	14.54	LSTM unggul tipis
$PM_{2.5}$	CNN-LSTM	9.78	12.58	15.63	CNN-LSTM unggul sangat tipis
SO_2	CNN-LSTM	1.70	2.70	5.77	CNN-LSTM jauh lebih baik
CO	LSTM	2.34	3.27	16.54	LSTM unggul
O_3	CNN-LSTM	3.09	3.99	13.63	CNN-LSTM terbaik
NO_2	CNN-LSTM	3.91	5.23	15.98	CNN-LSTM terbaik

Tabel 2 menunjukkan hasil evaluasi yang menyeluruh untuk semua polutan, Model LSTM performa yang sangat baik, terutama pada $PM_{2.5}$, PM_{10} dan CO. Hal ini menunjukkan LSTM mampu menangkap pola temporal dengan baik. Sementara secara keseluruhan CNN-LSTM merupakan model yang paling unggul. Keunggulan CNN-LSTM terbesar terlihat pada polutan NO_2 , SO_2 , dan O_3 . Kombinasi kemampuan CNN menangkap pola spasial ditambah LSTM menangkap urutan waktu membuat model ini paling efektif untuk data polusi udara kompleks.

Gambar 7 memperlihatkan perbandingan antara data aktual $PM_{2.5}$, hasil prediksi historis, dan hasil peramalan 30 hari ke depan yang dihasilkan oleh model ARIMA, LSTM, CNN-LSTM, dan *hybrid* ARIMA-LSTM. Visualisasi ini digunakan untuk mengevaluasi kemampuan masing-masing model dalam mengikuti pola historis data serta memproyeksikan kondisi kualitas udara pada periode mendatang.

Pada Gambar 7 garis berwarna biru merepresentasikan data aktual $PM_{2.5}$ yang menunjukkan pola fluktuatif sepanjang periode pengamatan. Terlihat bahwa konsentrasi $PM_{2.5}$ mengalami perubahan yang cukup signifikan dari waktu ke waktu, dengan beberapa lonjakan pada periode tertentu yang mengindikasikan peningkatan pencemaran udara. Pola ini menunjukkan bahwa data kualitas udara memiliki karakteristik non-linear dan volatilitas tinggi.



Gambar 7. Perbandingan Hasil Prediksi Model dengan Data Aktual dan Peramalan

Temuan pada data historis mengonfirmasi bahwa arsitektur *deep learning*, terutama varian CNN-LSTM dan LSTM, memiliki performa yang lebih unggul dalam mereplikasi pola data riil jika dibandingkan dengan model statistik ARIMA. Garis prediksi CNN-LSTM terlihat lebih dekat dengan data aktual, terutama pada periode ketika terjadi perubahan konsentrasi $PM_{2.5}$ secara cepat. Hal ini menegaskan bahwa penggabungan CNN dan LSTM memberikan keunggulan dalam mengidentifikasi pola-pola rumit, baik dari sisi fluktuasi waktu maupun fitur-fitur spesifik pada data kualitas udara.

Model ARIMA cenderung memproduksi estimasi yang lebih halus (*smoother*) serta kurang responsif terhadap lonjakan data sebenarnya. Kondisi ini membuktikan bahwa model statistik linear sulit mengadaptasi kompleksitas dan sifat non-linearitas dari data $PM_{2.5}$. Sementara itu, model *hybrid* ARIMA-LSTM menghasilkan performa yang lebih baik dibandingkan ARIMA, namun masih belum mampu melampaui performa CNN-LSTM dan LSTM. Pada bagian akhir grafik ditampilkan hasil *forecasting* selama 30 hari ke depan menggunakan masing-masing model. Meski secara umum pola $PM_{2.5}$ yang dihasilkan oleh seluruh model tampak stabil pada rentang waktu tertentu, CNN-LSTM terbukti lebih responsif dan adaptif dalam menangkap dinamika data. Hal ini mengindikasikan bahwa model CNN-LSTM memiliki kemampuan yang lebih baik dalam mempertahankan pola temporal dari data historis untuk digunakan dalam prediksi jangka pendek.

Dari hasil visualisasi tersebut dapat disimpulkan bahwa model *deep learning*, terutama CNN-LSTM lebih efektif dalam memodelkan data kualitas udara perkotaan yang memiliki pola kompleks dan fluktuatif. Temuan ini memperkuat hasil evaluasi model sebelumnya yang menunjukkan bahwa CNN-LSTM memiliki nilai error paling rendah dibandingkan model lainnya. Selain itu, CNN-LSTM juga menunjukkan performa yang baik untuk semua polutan yang diteliti.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa model *deep learning* dan *hybrid deep learning* memiliki performa yang lebih baik dibandingkan model statistik tradisional dalam peramalan Indeks Standar Pencemar Udara (ISPU) di DKI Jakarta. Model CNN-LSTM menghasilkan performa terbaik dengan nilai MAE sebesar 4.54, RMSE sebesar 6.03, dan MAPE sebesar 13.83%, diikuti oleh LSTM yang menunjukkan hasil evaluasi yang kompetitif. Sementara itu, ARIMA menghasilkan performa terendah karena kurang mampu menangkap pola non-linear pada data kualitas udara perkotaan. Model *hybrid* ARIMA-LSTM juga belum mampu melampaui performa model *deep learning* tunggal, yang menunjukkan bahwa data ISPU DKI Jakarta lebih didominasi oleh pola non-linear. Hasil visualisasi prediksi memperlihatkan bahwa model CNN-LSTM mampu mengikuti fluktuasi $PM_{2.5}$ dengan lebih baik dibandingkan model lainnya dan hal ini juga berlaku untuk semua polutan yang diteliti dalam penelitian ini, sehingga berpotensi digunakan dalam sistem prediksi kualitas udara dan pengendalian pencemaran udara di wilayah perkotaan. Penelitian selanjutnya disarankan menambahkan variabel meteorologi, mengembangkan model *deep learning* yang lebih kompleks seperti *Transformer* atau *BiLSTM*, serta menerapkan pendekatan SHAP agar interpretasi faktor dominan pencemaran udara dapat dianalisis secara lebih mendalam dan mendukung pengambilan kebijakan lingkungan yang lebih efektif.

DAFTAR PUSTAKA

- [1] R. Wangintan, M. Yani, H. Hardjomidjojo, and T. June, "Evaluation of Air Pollution Control Management in Jakarta by Crossed-Impact Matrix Multiplications Applied to Classification (MICMAC)," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1358, no. 1, 2024, doi: 10.1088/1755-1315/1358/1/012044.
- [2] Farhatun Haya, Khaira Nisa, Rio Febrian Ladipasa, Ari Suriani, and Afriza Media, "Dampak Polusi Udara terhadap Kesehatan Manusia," *WISSEN: Jurnal Ilmu Sosial dan Humaniora*, vol. 3, no. 2, pp. 180–190, 2025, doi: 10.62383/wissen.v3i2.753.
- [3] B. -, S. Listyarini, and L. Warlina, "Pengaruh Kebijakan Pencemaran Udara Sektor Transportasi Terhadap Nilai Indeks Kualitas Udara (Iku) Di Dki Jakarta," *Jurnal Ilmiah Pendidikan Lingkungan dan Pembangunan*, vol. 24, no. 01, pp. 1–13, 2023, doi: 10.21009/plpb.v24i01.30798.
- [4] M. Mushliha, "Implementasi CNN-BiLSTM untuk Prediksi Harga Saham Bank Syariah di Indonesia," *Jambura Journal of Mathematics*, vol. 6, no. 2, pp. 195–203, 2024, doi: 10.37905/jjom.v6i2.26509.
- [5] Z. Shahbazi, Z. Shahbazi, and S. Nowaczyk, "Enhancing Air Quality Forecasting Using Machine Learning Techniques," *IEEE Access*, vol. 12, no. December, pp. 197290–197299, 2024, doi: 10.1109/ACCESS.2024.3516883.
- [6] S. Deepan and M. Saravanan, "Air quality index prediction using seasonal autoregressive integrated moving average transductive long short-term memory," *ETRI Journal*, vol. 46, no. 5, pp. 915–927, 2024, doi: 10.4218/etrij.2023-0283.
- [7] K. Sekaran, M. Priyadharshini, and R. Sahay, "AirTFT: A Novel Transformer-Based Approach for Air Quality Prediction," *International Symposium on Advanced Networks and Telecommunication Systems, ANTS*, pp. 114–119, 2024, doi: 10.1109/ANTS63515.2024.10898497.
- [8] S. Tilahun, K. Abera, A. Habtie, B. Tilahun, M. Gebrehiwot, and T. N. Mengestie, "Predicting Air Quality for Sustainable Urban Living," *2025 International Conference on Information and Communication Technology for Development for Africa (ICT4DA)*, no. ML, pp. 55–59, 2025, doi: 10.1109/ict4da67218.2025.11282769.

- [9] J. Xiao, Q. Wang, J. Cui, and J. Yu, "Multi-feature PM2.5 Prediction with ARIMA-LSTM," *Proceedings - 2022 International Conference on Machine Learning, Cloud Computing and Intelligent Mining, MLCCIM 2022*, no. Mlccim, pp. 229–233, 2022, doi: 10.1109/MLCCIM55934.2022.00046.
- [10] K. Kashif and R. Ślepaczuk, "LSTM-ARIMA as a hybrid approach in algorithmic investment strategies," *Knowl. Based. Syst.*, vol. 320, no. March, 2025, doi: 10.1016/j.knosys.2025.113563.
- [11] J. Zhang and S. Li, "Air quality index forecast in Beijing based on CNN-LSTM multi-model," *Chemosphere*, vol. 308, no. P1, p. 136180, 2022, doi: 10.1016/j.chemosphere.2022.136180.
- [12] C. Shilengwe, K. Banda, and I. Nyambe, "Machine learning downscaling of GRACE/GRACE-FO data to capture spatial-temporal drought effects on groundwater storage at a local scale under data-scarcity," *Environmental Systems Research*, vol. 13, no. 1, pp. 1–19, 2024, doi: 10.1186/s40068-024-00368-1.
- [13] D. L. John, S. Binnewies, and B. Stantic, "Identifying Optimal Window Size Configurations for Big Data Time Series Forecasting," *Proceedings - 2024 IEEE International Conference on Big Data, BigData 2024*, pp. 5138–5146, 2024, doi: 10.1109/BigData62323.2024.10825222.
- [14] P. A. Faradiela, A. Wibowo, and Y. R. Husniyawati, "Forecasting the Number of Vertical Referrals for Bpjs Participants At Health Service Center in Universitas Airlangga Using Arima Model," *Jurnal Biometrika dan Kependudukan*, vol. 11, no. 1, pp. 62–71, 2022, doi: 10.20473/jbk.v11i1.2022.62-71.
- [15] F. A. Mahbubi, T. I. Hermanto, and C. D. Lestari, "Peramalan Penjualan Saham Nikel Menggunakan Algoritma Long Short Term Memory (Lstm)," *IDEALIS: InDonEsiA journal Information System*, vol. 8, no. 1, pp. 138–149, 2025, doi: 10.36080/idealis.v8i1.3254.
- [16] F. Aksan, Y. Li, V. Suresh, and P. Janik, "CNN-LSTM vs. LSTM-CNN to Predict Power Flow Direction: A Case Study of the High-Voltage Subnet of Northeast Germany," *Sensors*, vol. 23, no. 2, 2023, doi: 10.3390/s23020901.
- [17] T. A. Wahyudi *et al.*, "Pengembangan E - Marketplace Ikan dengan Chatbot NLP berbasis LSTM di BBI Cangkiran Mijen," vol. 14, pp. 2299–2317, 2025, doi: <https://doi.org/10.32520/stmsi.v14i5.5170>.
- [18] C. Li, X. Fang, Z. Yan, Y. Huang, and M. Liang, "Research on Gas Concentration Prediction Based on the ARIMA-LSTM Combination Model," *Processes*, vol. 11, no. 1, 2023, doi: 10.3390/pr11010174.
- [19] R. Dadi, C. Narsingoju, K. Sridhar, K. Sravani, V. Mounika, and Shabana, "Temporal Fusion and Hybrid ARIMA-LSTM Models for Robust Time Series Forecasting: A Case Study on Ethereum Price Dynamics," *Proceedings of 2025 2nd International Conference on Cognitive Robotics and Intelligent Systems, ICC - ROBINS 2025*, pp. 622–627, 2025, doi: 10.1109/ICC-ROBINS64345.2025.11086123.
- [20] B. Bhagawanta and C. B. Santoso, "Deteksi penyakit daun kapas dengan deep learning berbasis Convolutional Neural Network (CNN)," *IDEALIS InDonEsiA journal Information System*, vol. 8, no. 2, pp. 220–227, Jul. 2025, doi: 10.36080/idealis.v8i2.3517.
- [21] Y. Ding, S. Liu, Z. Zhao, and S. Liu, "A CNN-LSTM Model for Electric Load Forecasting," *ICNSC 2024 - 21st International Conference on Networking, Sensing and Control: Artificial Intelligence for the Next Industrial Revolution*, pp. 1–6, 2024, doi: 10.1109/ICNSC62968.2024.10760118.