

ANALISIS PREDIKSI EMISI KARBON DIOKSIDA KENDARAAN BERMOTOR MENGGUNAKAN ALGORITMA HIBRIDA LSTM DAN ARIMA

Muhammad Hakam Fardana^{1*}, Wahyu Syaifullah Jauharis Saputra, S.Kom., M.Kom.², Made Hanindia Prami S, S.Kom., M.Cs.³

^{1,2,3}Informatika, Universitas Pembangunan Nasional Veteran Jawa Timur; Jalan Rungkut Madya No. 1 Gunung Anyar, Surabaya, Indonesia

Keywords:

Emisi CO₂ Kendaraan
Bermotor; Algoritma Hibrida
LSTM dan ARIMA;
Peramalan

Correspondent Email:

mhakamf13@gmail.com

Abstrak. Emisi CO₂ yang dihasilkan oleh kendaraan bermotor memberikan kontribusi signifikan terhadap perubahan iklim global. Oleh karena itu, peramalan yang akurat terhadap tren emisi menjadi sangat penting dalam upaya mitigasi. Penelitian ini mengkaji kinerja model hibrida antara Long Short-Term Memory (LSTM) dan Autoregressive Integrated Moving Average (ARIMA) dalam memprediksi emisi CO₂ dari kendaraan bermotor. Pendekatan hibrida ini memadukan keunggulan ARIMA dalam menangkap pola linier serta kemampuan LSTM dalam mengenali pola non-linier jangka panjang. Dengan memanfaatkan 1000 data historis yang diperoleh dari Aplikasi Eco-Route, model ini diuji dan dibandingkan dengan model individual. Hasil pengujian menunjukkan bahwa model hibrida mampu memberikan akurasi yang baik, dengan nilai MAE sebesar 0,0941, MAPE 10,20%, dan RMSE 0,1081 pada skenario terbaik. Namun demikian, model ARIMA tunggal memberikan hasil prediksi paling akurat dengan MAE 0,0835, MAPE 9,33%, dan RMSE 0,0975. Keterbatasan jumlah dan cakupan data mempengaruhi performa model hibrida. Dengan demikian, model LSTM-ARIMA dapat dianggap sebagai pendekatan alternatif yang menjanjikan dalam peramalan emisi CO₂, khususnya apabila diterapkan pada dataset yang lebih luas dan representatif.

Abstract. Motor vehicle CO₂ emissions significantly contribute to climate change, making accurate forecasting of emission trends essential for effective mitigation strategies. This study investigates the effectiveness of a hybrid model combining Long Short-Term Memory (LSTM) and Autoregressive Integrated Moving Average (ARIMA) in predicting CO₂ emissions from motor vehicles. The hybrid approach leverages ARIMA's strength in modeling linear trends and LSTM's proficiency in capturing long-term, non-linear relationships. Utilizing 1,000 historical records sourced from the Eco-Route Application, the hybrid model's performance was assessed and compared to that of individual models. The findings indicate that the hybrid model achieved commendable accuracy, with MAE of 0.0941, MAPE of 10.20%, and RMSE of 0.1081 under optimal conditions. Nevertheless, the standalone ARIMA model outperformed the others on this dataset, achieving MAE of 0.0835, MAPE of 9.33%, and RMSE of 0.0975. The limited dataset was identified as a factor affecting the hybrid model's performance. In conclusion, the Hybrid LSTM-ARIMA model presents a promising alternative for CO₂ emission prediction, particularly when applied to more comprehensive datasets.

1. PENDAHULUAN

Peningkatan emisi karbon dioksida (CO₂) yang signifikan dari kendaraan bermotor merupakan tantangan lingkungan mendesak yang berkontribusi besar pada pemanasan global. Dalam konteks ini, kemajuan teknologi di era digital memungkinkan pengembangan solusi inovatif, khususnya melalui teknologi prediktif. Untuk analisis dan peramalan tren emisi CO₂ secara akurat, algoritma pembelajaran mesin seperti Long Short-Term Memory (LSTM) dan Autoregressive Integrated Moving Average (ARIMA) menjadi semakin krusial. Penelitian ini bertujuan mengevaluasi performa dan akurasi prediksi emisi CO₂ dengan mengintegrasikan kedua metode tersebut ke dalam model hibrida. Temuan studi ini diharapkan dapat menjadi acuan dalam perumusan kebijakan pengurangan emisi yang lebih efektif dan efisien di masa mendatang, serta berkontribusi pada upaya pelestarian lingkungan global. Di era globalisasi ekonomi, masalah Emisi CO₂ dan perubahan iklim telah muncul dan menarik perhatian para sarjana [1]. Menurut Intergovernmental Panel on Climate Change (IPCC), jumlah CO₂ di atmosfer meningkat dengan kecepatan 2% per tahun. Pada tahun 2014, Emisi CO₂ global mencapai 36,14 miliar ton, tiga kali lipat dari jumlah yang dihasilkan pada tahun 1960. Emisi CO₂ yang besar ini memperburuk proses pemanasan global dan tindakan yang kuat diperlukan untuk menjaga pembangunan global yang berkelanjutan [2].

Tren kenaikan emisi Gas Rumah Kaca atau GRK, menaik sekitar 4,3% dari 2010 hingga 2018 yang menjadikan sektor transportasi sebagai salah satu penghasil emisi karbon terbesar dengan persentase 96,36% atau 28.317 ton per tahun [3]. Karena konsekuensi negatif yang signifikan dari perubahan iklim, masalah lingkungan yang disebabkan oleh transportasi darat semakin memburuk seiring dengan percepatan urbanisasi. Sementara permintaan transportasi meningkat, sektor transportasi harus menghadapi kesulitan dalam mengurangi emisi [4]. Jumlah gas rumah kaca yang dilepaskan ke lingkungan telah turun secara signifikan akibat pandemi COVID-19 pada tahun 2020, namun suhu global masih terus meningkat sebagai akibat dari peningkatan tingkat polusi udara [5]. Sebagai salah satu arsitektur Jaringan Saraf Tiruan (JST),

algoritma Long Short-Term Memory (LSTM) sangat efektif dalam memproses data deret waktu berkat kemampuannya mengingat pola dan informasi jangka panjang. Kemampuan LSTM untuk menangkap dependensi jangka panjang dalam data deret waktu menjadikannya krusial untuk menganalisis dan memahami tren Emisi CO₂ Kendaraan Bermotor. Di sisi lain, model statistik Autoregressive Integrated Moving Average (ARIMA) lazim digunakan untuk analisis dan peramalan data deret waktu. ARIMA terdiri dari elemen Autoregressive (AR), Moving Average (MA), dan operasi differencing yang memungkinkan penyesuaian terhadap tren serta pola musiman data, sehingga berkontribusi dalam mengestimasi fluktuasi Emisi CO₂ Kendaraan Bermotor secara efektif. Urbanisasi yang pesat dan meningkatnya mobilitas mendorong pertumbuhan masyarakat jumlah kendaraan bermotor, yang berkontribusi terhadap peningkatan Emisi CO₂. Sekitar 75% dari total Emisi CO₂ sektor transportasi global berasal dari kendaraan penumpang, menunjukkan besarnya kontribusi kendaraan bermotor dalam akumulasi karbon [6]. Untuk mendukung upaya mitigasi, diperlukan pemahaman yang komprehensif terhadap pola emisi serta kemampuan dalam memprediksi tren di masa depan. Analisis yang akurat dapat menjadi landasan ilmiah bagi perumusan kebijakan yang terukur serta pengembangan strategi pengurangan emisi yang lebih tepat sasaran.

Penelitian ini bertujuan utama menganalisis dan memproyeksikan tren emisi CO₂ dari kendaraan bermotor di Indonesia melalui pengembangan model deret waktu. Dengan mengintegrasikan algoritma hibrida Long Short-Term Memory (LSTM) dan Autoregressive Integrated Moving Average (ARIMA), studi ini berupaya meningkatkan akurasi prediksi. Peningkatan akurasi ini diharapkan memberikan kontribusi signifikan dalam pengambilan keputusan terkait pengelolaan emisi dan perumusan kebijakan lingkungan yang lebih efektif. Selain itu, penelitian ini mengeksplorasi potensi penuh dari algoritma hibrida LSTM dan ARIMA untuk analisis dan prediksi emisi CO₂, yang hasilnya dapat menjadi landasan berharga bagi pengembangan strategi mitigasi emisi yang efektif dan berkelanjutan di sektor transportasi. Prediksi yang akurat juga diharapkan dapat

meningkatkan kesadaran dan mendorong tindakan nyata oleh berbagai pihak demi keberlanjutan lingkungan dan pengurangan dampak negatif emisi CO₂.

2. TINJAUAN PUSTAKA

2.1. Emisi Karbon Dioksida

Berdasarkan Peraturan Pemerintah Republik Indonesia Nomor 41 Tahun 1999 mengenai Pengendalian Pencemaran Udara, emisi karbon diartikan sebagai zat, energi, dan/atau unsur lain yang berasal dari suatu aktivitas dan dilepaskan ke udara ambien, baik yang berpotensi maupun tidak berpotensi mencemari udara. Emisi karbon merujuk pada jumlah gas CO₂ yang dihasilkan, baik secara langsung maupun tidak langsung, dari aktivitas manusia. Umumnya, emisi ini dinyatakan dalam satuan ton karbon dioksida [7].

2.2. Kendaraan Bermotor

Kendaraan bermotor adalah kendaraan yang digunakan untuk transportasi darat dan digerakkan oleh peralatan teknik untuk pergerakannya. Perpindahan orang atau barang dari satu tempat ke tempat yang lain disebut transportasi. Transportasi atau pengangkutan adalah perpindahan dari satu tempat ke tempat lain dengan menggunakan alat transportasi, baik yang digerakkan oleh manusia, hewan (kuda, sapi, atau kerbau), atau mesin [8]. Oleh karena itu, transportasi jalan dapat dibagi menjadi dua kelompok besar: kendaraan bermotor dan kendaraan tidak bermotor.

2.3. Machine Learning

Algoritma Machine Learning merupakan suatu proses komputasional yang memanfaatkan data masukan untuk menyelesaikan tugas tertentu tanpa harus diprogram secara eksplisit (dengan "kode keras") guna menghasilkan keluaran yang diharapkan. Algoritma ini bersifat "soft coded" karena mampu menyesuaikan atau memodifikasi struktur internalnya secara otomatis melalui proses berulang (yakni, berdasarkan pengalaman), sehingga kinerjanya dalam menyelesaikan tugas yang dimaksud semakin meningkat. Proses penyesuaian ini dikenal dengan istilah pelatihan, dan dilakukan menggunakan sampel data masukan.

2.4. Forecasting

Peramalan, atau yang dikenal juga dengan istilah *forecasting*, merupakan proses

memperkirakan kejadian di masa mendatang dengan memanfaatkan data historis yang relevan dan menerapkannya melalui model matematis [9]. Peramalan dapat berfungsi sebagai dasar dalam penyusunan rencana jangka pendek, menengah, maupun panjang [10]. Secara umum, peramalan adalah suatu proses memprediksi nilai atau kejadian di masa depan berdasarkan pola data masa lalu [11]. Di sisi lain, prediksi lebih menekankan pada penilaian intuitif terhadap kejadian di masa depan yang tidak sepenuhnya bergantung pada pengalaman atau data sebelumnya [12]. Dengan demikian, *forecasting* dapat diartikan sebagai metode analitis yang menggabungkan pendekatan kuantitatif dan kualitatif untuk memperkirakan kejadian di masa depan dengan mempertimbangkan informasi historis yang tersedia.

2.5. Time Series

Analisis data deret waktu merupakan tahapan krusial dalam mengidentifikasi pola dan kecenderungan dalam data yang bersifat temporal [13]. Peramalan deret waktu sendiri merupakan metode yang digunakan untuk memperkirakan kejadian di masa depan berdasarkan urutan waktu. Teknik ini dilakukan dengan menganalisis pola historis dan tren masa lalu, dengan asumsi bahwa pola tersebut akan berlanjut secara serupa di masa yang akan datang [14].

2.6. Long Short-Term Memory (LSTM)

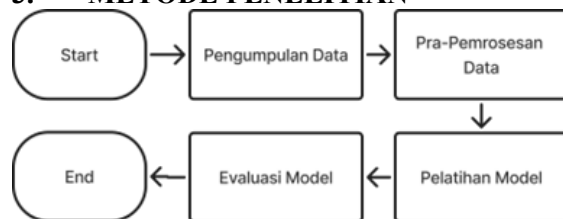
Long Short-Term Memory (LSTM) merupakan pengembangan dari arsitektur Recurrent Neural Network (RNN) yang dirancang untuk mengatasi permasalahan *vanishing gradient* serta mampu menangkap dependensi jangka panjang dalam data sekuensial melalui mekanisme *memori sel*, yang memungkinkan penyimpanan informasi dalam periode waktu yang lebih lama. LSTM hadir sebagai solusi terhadap keterbatasan RNN dalam memproses data dengan rentang waktu yang panjang [15]. Dengan kemampuannya dalam mempertahankan informasi historis yang relevan, LSTM dinilai lebih unggul dibandingkan algoritma lain dalam menangani data deret waktu [16].

2.7. Autoregressive Integrated Moving Average (ARIMA)

ARIMA (Autoregressive Integrated Moving Average) merupakan salah satu

algoritma yang umum digunakan dalam peramalan data deret waktu. Model ini dikenal andal dalam memodelkan data univariat, khususnya yang memperlihatkan pola linier. Meski demikian, ARIMA memiliki keterbatasan karena hanya mampu menangani satu variabel deret waktu, sehingga akurasi dapat menurun apabila data dipengaruhi oleh variabel eksternal atau mengandung pola non-linier yang tidak terwakili dalam observasi utama [17]. ARIMA sendiri merupakan pengembangan dari model Autoregressive Moving Average (ARMA), yang dirancang untuk menangani data non-stasioner dengan menerapkan proses integrasi guna menstabilkan rata-rata dan varians data sebelum pemodelan dilakukan [18].

3. METODE PENELITIAN



Gambar 1. Alur Penelitian

3.1. Pengumpulan Data

Pengumpulan data historis emisi karbon dioksida (CO₂) kendaraan bermotor yang mencakup fitur-fitur relevan merupakan langkah fundamental dalam penelitian ini. Dataset yang digunakan merepresentasikan total emisi CO₂ harian yang dihasilkan oleh kendaraan bermotor dengan spesifikasi seperti ukuran mesin 250cc, tahun produksi 2019, dan penggunaan bahan bakar Pertamina. Meskipun dataset awal mencakup beberapa fitur data seperti *time_used*, *emission(kg)*, *distance(km)*, *time_travel(h)*, dan *time_travel(m)*, proses analisis dan pemodelan dalam studi ini hanya memfokuskan pada variabel *time_used* dan *emission(kg)*. Fokus pada kedua variabel ini bertujuan untuk menyederhanakan proses analisis data dan menekankan pada fitur-fitur yang umumnya digunakan dalam literatur prediksi emisi CO₂.

3.2. Pra-Pemrosesan Data

Pra-pemrosesan data merupakan tahap fundamental dalam alur analisis data. Fungsi utamanya adalah membersihkan,

mengorganisasi, dan mempersiapkan data mentah, yang krusial untuk memastikan kualitas data dan menjamin akurasi output dari pemodelan atau analisis lanjutan. Tahap ini sangat penting untuk mengidentifikasi dan menangani berbagai isu data, termasuk data yang hilang, nilai outlier, dan noise, yang berpotensi signifikan mempengaruhi performa model. Lebih lanjut, pra-pemrosesan data berkontribusi dalam meningkatkan efisiensi algoritma pembelajaran mesin dalam menginterpretasikan data secara efektif.

3.3. Pelatihan Model

3.3.1. LSTM



Gambar 2. Alur Penelitian LSTM

a. Validasi Fitur

Setelah data berhasil dikumpulkan, tahap selanjutnya adalah menentukan fitur-fitur spesifik dari dataset yang akan digunakan dalam pemodelan. Untuk data yang akan diinput ke dalam model, dipilih dua fitur utama, yaitu *time_used* dan *emission(kg)*. Pemilihan kedua fitur ini didasarkan pada pertimbangan untuk memfokuskan ruang lingkup penelitian pada variabel kunci yang relevan dan untuk meningkatkan efisiensi dalam proses analisis serta pemodelan selanjutnya.

b. Pembentukan Sekuens

Dalam pemrosesan data deret waktu untuk model LSTM, tahap pembentukan sekuens (sequence formation) merupakan langkah krusial. Proses ini melibatkan restrukturisasi dataset terkumpul menjadi format sekuensial agar model dapat memahami dan mempelajari pola temporal data. Data disusun menjadi sekuens dengan menentukan ukuran window (window size), yaitu jumlah data poin yang digunakan untuk memprediksi data poin berikutnya.

c. Pembentukan Model LSTM

Pembentukan model LSTM adalah fase esensial untuk prediksi deret waktu akurat. Tahap ini mencakup konfigurasi hyperparameter kunci seperti learning rate,

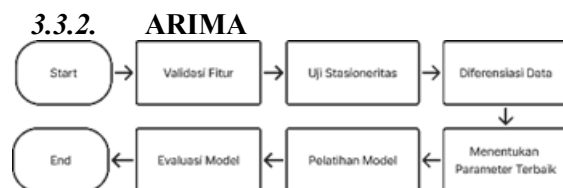
epoch, dan batch size yang mempengaruhi proses pembelajaran model.

d. Pelatihan Model

Setelah dibentuk, model dilatih menggunakan data historis; proses ini, di mana model mempelajari pola prediksi, sangat bergantung pada arsitektur dan konfigurasi awal yang telah ditetapkan.

e. Prediksi dan Evaluasi Model

Setelah pelatihan selesai, prediksi dilakukan terhadap data uji (test data) menggunakan model terbaik yang telah dipilih untuk masing-masing skenario pembagian data. Kinerja model dievaluasi menggunakan metrik Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), dan Root Mean Square Error (RMSE). Hasil evaluasi kinerja model LSTM ini kemudian secara komparatif dianalisis dengan performa model ARIMA dan model Hibrida LSTM-ARIMA guna menilai efektivitasnya dalam prediksi total Emisi CO₂.



Gambar 3. Alur Penelitian ARIMA

a. Validasi Fitur

Sebagai model univariat, pemodelan Autoregressive Integrated Moving Average (ARIMA) memerlukan satu variabel deret waktu sebagai input utama. Dalam penelitian ini, variabel emission(kg) dipilih sebagai deret waktu target untuk dianalisis dan diramalkan menggunakan ARIMA.

b. Uji Stasioneritas

Jika data belum mencapai kondisi stasioner, diperlukan transformasi seperti differencing atau logaritmik untuk menstabilkan pola sebelum diterapkan ke model ARIMA. Setelah transformasi dilakukan, uji ADF kembali diterapkan guna memastikan data telah memenuhi karakteristik stasioner. Langkah ini krusial karena model ARIMA bekerja lebih efektif pada data yang stabil, sehingga prediksi yang dihasilkan menjadi lebih akurat dan dapat diandalkan.

c. Diferensiasi Data

Validasi stasioneritas deret waktu merupakan langkah awal yang krusial. Jika data belum menunjukkan karakteristik stasioner, transformasi seperti differencing atau logaritmik diterapkan untuk menstabilkan polanya. Uji Augmented Dickey-Fuller (ADF) kemudian dilakukan untuk memverifikasi stasioneritas data setelah transformasi. Proses differencing secara spesifik menghitung selisih antar observasi berurutan dan dapat diulang sesuai kebutuhan hingga data mencapai kondisi stasioner, yang esensial untuk efektivitas model ARIMA.

d. Menemukan Parameter Terbaik

Setelah data terkonfirmasi stasioner, langkah berikutnya adalah menentukan parameter model ARIMA (p,d,q). Nilai parameter d diperoleh langsung dari jumlah differencing yang dibutuhkan untuk mencapai stasioneritas. Sementara itu, parameter p (Autoregressive) dan q (Moving Average) diidentifikasi melalui analisis visual plot Autocorrelation Function (ACF) dan Partial Autocorrelation Function (PACF), dengan p berdasarkan lag signifikan di plot PACF dan q berdasarkan lag signifikan di plot ACF.

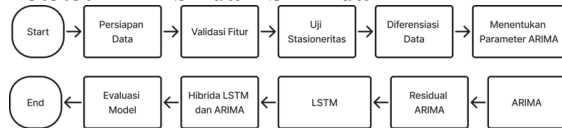
Pemilihan model ARIMA terbaik dari berbagai kombinasi parameter dilakukan menggunakan Akaike Information Criterion (AIC). Model dengan nilai AIC terendah dipilih, karena metrik ini mengindikasikan keseimbangan optimal antara akurasi prediksi model dan kompleksitasnya, sehingga cocok untuk analisis dan peramalan deret waktu.

e. Pelatihan Model

Setelah pelatihan selesai, prediksi dilakukan terhadap data uji (test data) menggunakan model ARIMA terbaik yang telah diperoleh. Prediksi ini dilaksanakan untuk setiap skenario pembagian data yang digunakan dalam penelitian. Kinerja prediksi model dievaluasi menggunakan metrik Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), dan Root Mean Square Error (RMSE). Hasil evaluasi kinerja model ARIMA ini kemudian dibandingkan secara komparatif dengan performa model LSTM dan model Hibrida LSTM-ARIMA untuk menilai

efektivitas relatifnya dalam memprediksi total Emisi CO₂.

3.3.3. Hibrida LSTM dan ARIMA



Gambar 4. Alur Penelitian Hibrida LSTM dan ARIMA

a. Persiapan Data

Langkah pertama dimulai dengan melakukan persiapan terhadap dataset yang akan digunakan. Proses ini mencakup pemuatan data historis dari sumber yang tersedia ke dalam lingkungan pemrograman, serta pembersihan data melalui penghapusan nilai-nilai yang hilang (missing values), penanganan duplikasi, dan transformasi data jika diperlukan. Selain itu, data diurutkan berdasarkan waktu (timestamp) guna memastikan bahwa urutan kronologis data tetap terjaga. Tahap ini sangat krusial karena kualitas data yang digunakan akan sangat mempengaruhi kinerja model yang dikembangkan.

b. Validasi Fitur

Setelah data siap digunakan, dilakukan validasi terhadap fitur yang menjadi fokus penelitian, yaitu emisi(kg). Validasi ini bertujuan untuk memastikan bahwa fitur tersebut bersifat numerik, kontinu, serta representatif terhadap tren emisi karbon harian kendaraan. Selain itu, validasi juga bertujuan untuk mengidentifikasi adanya pola outlier, fluktuasi ekstrem, serta hubungan korelatif dengan waktu yang akan berdampak terhadap stasioneritas dan efektivitas pemodelan.

c. Uji Stasioneritas

Model ARIMA memiliki asumsi dasar bahwa data yang akan dimodelkan bersifat stasioner, yaitu memiliki rata-rata dan variansi yang konstan sepanjang waktu. Oleh karena itu, pada tahap ini dilakukan pengujian stasioneritas terhadap data target menggunakan uji statistik ADF Test. Hasil dari pengujian ini digunakan sebagai dasar pengambilan keputusan apakah data perlu dilakukan transformasi lanjutan seperti diferensiasi.

d. Diferensiasi Data

Apabila data dinyatakan tidak stasioner berdasarkan hasil uji ADF, maka langkah selanjutnya adalah melakukan proses diferensiasi data (differencing). Diferensiasi bertujuan untuk menghilangkan tren dalam data sehingga pola pergerakan data lebih stabil. Proses ini dapat dilakukan satu kali (first-order differencing) atau lebih, bergantung pada tingkat ketidakstasioneran yang terdeteksi. Hasil dari diferensiasi akan digunakan sebagai input utama dalam pelatihan model ARIMA.

e. Penentuan Parameter ARIMA

Model ARIMA memiliki tiga parameter utama, yaitu p (order AR), d (order differencing), dan q (order MA). Untuk menentukan kombinasi parameter yang optimal, dilakukan pendekatan berbasis analisis Autocorrelation Function (ACF) dan Partial Autocorrelation Function (PACF), atau menggunakan metode otomatis seperti Auto ARIMA. Tujuan dari tahap ini adalah untuk memperoleh parameter yang menghasilkan performa terbaik, biasanya ditentukan melalui nilai kriteria informasi seperti Akaike Information Criterion (AIC) dan Bayesian Information Criterion (BIC).

f. Pelatihan Model ARIMA

Pada tahap ini, model ARIMA dilatih menggunakan data historis yang telah melalui proses transformasi stasioneritas. Model ini bertugas menangkap pola linier dalam data emisi karbon. Setelah pelatihan selesai, model digunakan untuk melakukan prediksi nilai emisi karbon serta menghasilkan residual, yaitu selisih antara prediksi ARIMA dengan nilai aktual. Residual ini kemudian menjadi masukan penting dalam pengembangan model LSTM pada tahap selanjutnya.

g. Perhitungan Residual ARIMA

Hasil residual dari model ARIMA diasumsikan mengandung pola-pola non-linier yang tidak mampu dijelaskan oleh model linier. Oleh karena itu, residual ini kemudian diolah kembali dan digunakan sebagai input untuk pelatihan model LSTM. Tahap ini menjadi titik integrasi

antara kedua model, di mana LSTM akan berperan sebagai pemodel terhadap error yang ditinggalkan oleh ARIMA.

h. Pelatihan Model LSTM

Model LSTM dilatih menggunakan residual yang telah diproses sebelumnya, termasuk melalui tahap normalisasi dan pembentukan sekuens input sesuai dengan struktur time series. LSTM dipilih karena kemampuannya dalam mengenali pola dependensi jangka panjang serta menangani kompleksitas temporal yang tidak dapat ditangkap oleh model tradisional. Tujuan dari tahap ini adalah agar LSTM dapat mempelajari pola-pola dinamis yang tersisa dalam residual dan menghasilkan prediksi terhadap komponen non-linier dari data.

i. Pelatihan Model Hibrida LSTM dan ARIMA

Setelah kedua model (ARIMA dan LSTM) menghasilkan prediksi masing-masing, dilakukan proses integrasi untuk memperoleh hasil akhir dari model hybrid. Proses penggabungan ini dilakukan dengan cara menjumlahkan hasil prediksi ARIMA (komponen linier) dan prediksi LSTM terhadap residual (komponen non-linier). Hasil akhir dari proses ini merupakan prediksi yang menggabungkan kekuatan kedua model, sehingga diharapkan mampu memberikan hasil yang lebih akurat dan representatif terhadap pola data yang kompleks.

j. Prediksi dan Evaluasi Model

Setelah model selesai dilatih, prediksi dilakukan menggunakan data test dengan model terbaik yang telah dipilih. Prediksi dilakukan untuk masing-masing skenario pembagian data. Kinerja model dievaluasi menggunakan dua metrik utama, yaitu Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) dan Root Mean Square Error (RMSE). Hasil evaluasi dari model Hibrida LSTM dan ARIMA akan dibandingkan dengan metode lain seperti ARIMA dan LSTM untuk menilai performa model dalam prediksi total Emisi CO₂.

3.4. Evaluasi Model

Prediksi deret waktu secara inheren mengandung error dan tidak menghasilkan akurasi sempurna. Oleh karena itu, tahap

evaluasi krusial untuk mengukur dan memahami besarnya kesalahan prediksi. Dalam studi ini, evaluasi kinerja model dilaksanakan melalui serangkaian skenario pengujian komprehensif. Skenario ini melibatkan variasi rasio pembagian data pelatihan (train data) dan data pengujian (test data), serta pengaturan learning rate spesifik untuk model LSTM dan Hibrida LSTM-ARIMA. Rincian skenario pengujian adalah sebagai berikut

1. LSTM dan Hibrida LSTM dan ARIMA.
 - a. Variasi rasio data latih dan data uji
 - Skenario 1: 70% data latih dan 30% data uji.
 - Skenario 2: 80% data latih dan 20% data uji.
 - Skenario 3: 90 data latih dan 10% data uji.
 - b. Variasi nilai learning rate
 - 0.1
 - 0.01
 - 0.001
2. ARIMA
 - a. Skenario 1: 70% data latih dan 30% data uji.
 - b. Skenario 2: 80% data latih dan 20% data uji.
 - c. Skenario 3: 90 data latih dan 10% data uji.

Setiap skenario pengujian yang dilakukan akan dihitung errornya menggunakan MAE, MAPE, dan RMSE.

4. HASIL DAN PEMBAHASAN

Berikut ini adalah hasil penelitian yang didapatkan pada setiap alur penelitian yang dijabarkan pada bab sebelumnya.

4.1. Pengumpulan Data

Dataset historis emisi karbon dioksida (CO₂) dari kendaraan bermotor diakuisisi dan menjalani proses penyaringan. Tahap ini bertujuan untuk hanya menyisakan atribut tanggal pemakaian (time_used) dan total emisi (emission (kg)) yang relevan untuk analisis. Selanjutnya, data yang telah disaring kemudian dikonversi ke dalam format numerik. Konversi ini esensial guna memastikan kompatibilitas data dengan proses analisis dan pemodelan komputasional yang akan diterapkan.

	time_used	emission(kg)
0	2023-01-01	4876,1737
1	2023-01-02	5171,9746
2	2023-01-03	5585,0127
3	2023-01-04	6287,5015
4	2023-01-05	6148,1791

Gambar 5. Pengumpulan Data

4.2. Pelatihan Model

4.2.1. LSTM

a. Normalisasi Data

Untuk memastikan proses pelatihan yang optimal dan efisien, data total Emisi CO₂ dinormalisasi menggunakan metode Min-Max Scaling. Teknik ini mentransformasi data ke dalam rentang [0, 1] dan dipilih karena kemampuannya dalam mempertahankan hubungan proporsional antar nilai tanpa distorsi distribusi data.

b. Pembagian Data

Selanjutnya, dataset dibagi menjadi data latih (training data) dan data uji (testing data) guna memungkinkan evaluasi kinerja model secara objektif pada data yang belum pernah dilihat sebelumnya.

c. Pembentukan Sekuens

Data kemudian diubah menjadi format sekuensial, esensial agar model LSTM dapat mempelajari pola temporal dari data historis untuk memprediksi tren berikutnya.

d. Pembentukan Model LSTM

Arsitektur model LSTM yang diimplementasikan terdiri dari lapisan input untuk menerima data sekuensial Emisi CO₂, satu lapisan tersembunyi dengan neuron beraktivasi ReLU, dan lapisan output tunggal dengan fungsi aktivasi linear untuk menghasilkan prediksi total Emisi CO₂.

e. Pelatihan Model

Pelatihan model dilakukan menggunakan data latih dengan konfigurasi epochs 100 dan batch size 16. Proses pelatihan ini dieksplorasi melalui serangkaian skenario pengujian komprehensif, mencakup variasi rasio pembagian data latih:data uji (70:30,

80:20, 90:10) serta pengujian dengan tiga nilai learning rate berbeda (0.1, 0.01, 0.001). Variasi ini bertujuan mengamati dampak hyperparameter dan ukuran dataset terhadap akurasi model dan memastikan pembelajaran optimal tanpa overfitting.

f. Prediksi Model

Setelah model berhasil dilatih, prediksi total Emisi CO₂ dilakukan menggunakan dataset uji yang telah disiapkan sebelumnya.

4.2.2. ARIMA

a. Uji Stasioneritas

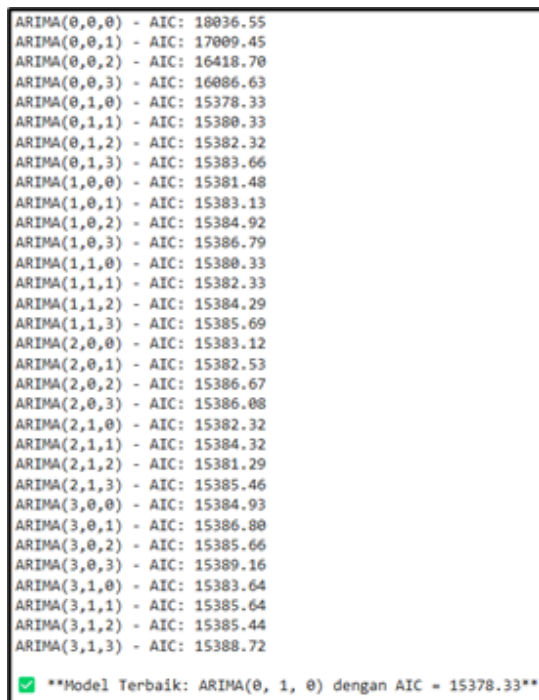
Stasioneritas data deret waktu merupakan syarat penting untuk pemodelan ARIMA. Kondisi ini diverifikasi menggunakan uji Augmented Dickey-Fuller (ADF). Data dianggap tidak stasioner apabila hasil uji memberikan nilai p-value lebih besar dari 0.05, mengindikasikan perlunya transformasi seperti differencing guna menstabilkan deret waktu.

```
Uji Stasioneritas:
ADF Statistic: -13.2791, p-value: 0.0000
Data sudah stasioner (p-value <= 0.05)
```

Gambar 6. Uji Stasioneritas

b. Penentuan Model ARIMA

Sebelum pelatihan model, dilakukan penentuan model ARIMA terbaik. Proses ini didasarkan pada pemilihan kombinasi parameter (p,d,q) yang menghasilkan nilai Akaike Information Criterion (AIC) terkecil. Nilai AIC yang rendah menunjukkan keseimbangan optimal antara kompleksitas model dan kecocokan data.



ARIMA(0,0,0) - AIC: 18036.55
ARIMA(0,0,1) - AIC: 17009.45
ARIMA(0,0,2) - AIC: 16418.70
ARIMA(0,0,3) - AIC: 16086.63
ARIMA(0,1,0) - AIC: 15378.33
ARIMA(0,1,1) - AIC: 15380.33
ARIMA(0,1,2) - AIC: 15382.32
ARIMA(0,1,3) - AIC: 15383.66
ARIMA(1,0,0) - AIC: 15381.48
ARIMA(1,0,1) - AIC: 15383.13
ARIMA(1,0,2) - AIC: 15384.92
ARIMA(1,0,3) - AIC: 15386.79
ARIMA(1,1,0) - AIC: 15380.33
ARIMA(1,1,1) - AIC: 15382.33
ARIMA(1,1,2) - AIC: 15384.29
ARIMA(1,1,3) - AIC: 15385.69
ARIMA(2,0,0) - AIC: 15383.12
ARIMA(2,0,1) - AIC: 15382.53
ARIMA(2,0,2) - AIC: 15386.67
ARIMA(2,0,3) - AIC: 15386.08
ARIMA(2,1,0) - AIC: 15382.32
ARIMA(2,1,1) - AIC: 15384.32
ARIMA(2,1,2) - AIC: 15381.29
ARIMA(2,1,3) - AIC: 15385.46
ARIMA(3,0,0) - AIC: 15384.93
ARIMA(3,0,1) - AIC: 15386.80
ARIMA(3,0,2) - AIC: 15385.66
ARIMA(3,0,3) - AIC: 15389.16
ARIMA(3,1,0) - AIC: 15383.64
ARIMA(3,1,1) - AIC: 15385.64
ARIMA(3,1,2) - AIC: 15385.44
ARIMA(3,1,3) - AIC: 15388.72

✓ **Model Terbaik: ARIMA(0, 1, 0) dengan AIC = 15378.33**

Gambar 7. Hasil Penentuan Model ARIMA berdasarkan AIC Terendah

c. Pembagian Data ARIMA

Setelah model ARIMA terbaik teridentifikasi dan data dipastikan stasioner (jika diperlukan differencing), data deret waktu dibagi menjadi data latih (training data) dan data uji (testing data) untuk keperluan pelatihan dan evaluasi independen.

d. Prediksi Model ARIMA

Menggunakan model ARIMA terbaik yang telah dilatih dengan data latih, prediksi Emisi CO₂ dilakukan terhadap data uji. Kinerja prediksi model dievaluasi dengan membandingkan hasil ramalan terhadap data aktual menggunakan metrik Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), dan Root Mean Square Error (RMSE). Selain itu, perbandingan visual antara data prediksi dan data aktual juga dilakukan melalui representasi grafis.

4.2.3. Hibrida LSTM dan ARIMA

a. Persiapan Data

Pada tahapan ini, data awal untuk model hibrida ini relatif serupa dengan prosedur untuk model individual. Dataset historis Emisi CO₂ dari kendaraan bermotor dimuat, disaring untuk hanya mencakup atribut `time_used` dan

`emission(kg)`, kemudian dikonversi ke dalam format numerik.

b. Pembagian Data ARIMA

Setelah proses penentuan parameter optimal ARIMA (p,d,q) dilakukan, data dibagi menjadi data latih dan data uji.

c. Pelatihan Model ARIMA

Sebagai langkah pertama dalam konstruksi model hibrida, model ARIMA yang telah dilatih digunakan untuk memprediksi Emisi CO₂ pada data uji. Hasil prediksi ARIMA ini kemudian dibandingkan dengan data aktual menggunakan metrik Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), dan Root Mean Square Error (RMSE), serta divisualisasikan melalui grafik. Residual dari prediksi ARIMA ini dihitung.

d. Perhitungan Residual ARIMA

Residual prediksi ARIMA diasumsikan memuat pola non-linier yang tidak sepenuhnya mampu ditangkap oleh struktur linier ARIMA. Oleh karena itu, residual ini diproses lebih lanjut dan dijadikan input bagi model LSTM.

e. Normalisasi Data Residual ARIMA

Data residual dinormalisasi dan diubah menjadi format sekuensial, sesuai dengan persyaratan input model LSTM.

f. Pembentukan Sekuens Residual

ARIMA untuk LSTM

Model LSTM dibangun untuk memprediksi residual, terdiri atas dua lapisan LSTM dan satu lapisan Dense sebagai output.

g. Pelatihan Residual ARIMA oleh LSTM

Proses pelatihan model residual menggunakan fungsi loss Mean Squared Error (MSE) dan optimizer Adam. Untuk optimasi performa, diuji tiga variasi nilai learning rate (0.1, 0.01, dan 0.001). LSTM bertugas memprediksi residual ini.

h. Integrasi Model Hibrida LSTM dan ARIMA

Setelah pelatihan model residual selesai, prediksi residual oleh LSTM dikembalikan ke skala semula menggunakan inverse transform dari MinMaxScaler. Prediksi akhir model hibrida (Hybrid Forecasting) diperoleh dengan menjumlahkan hasil prediksi

ARIMA dan hasil prediksi residual dari LSTM. Pendekatan hibrida ini menggabungkan kekuatan model ARIMA dalam menangkap pola linier dan stasioner dengan kemampuan LSTM menangkap pola non-linier dan noise residual, menghasilkan prediksi yang lebih komprehensif dan berpotensi lebih akurat dengan mengakomodasi karakteristik temporal data secara menyeluruh.

4.3. Evaluasi Model

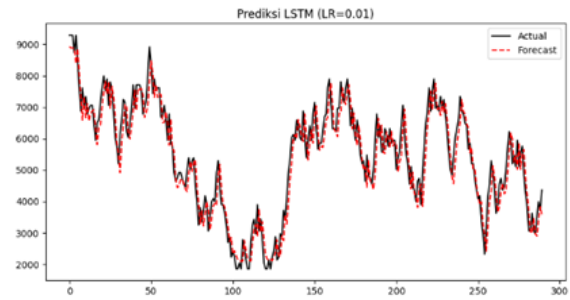
Setelah masing-masing model dilatih dengan skema pengujian yang sudah ditentukan sebelumnya, didapatkan hasil akhir dari setiap skenario pengujian.

4.3.1. LSTM

Tabel 1. Hasil Evaluasi Prediksi LSTM

Data Train (%)	Data Test (%)	Lr	MAE	MAPE	RMSE
70	30	0.1	0.1006	11.61%	0.1217
		0.01	0.0865	9.49%	0.1008
		0.001	0.0990	11.55%	0.1185
80	20	0.1	0.1008	10.75%	0.1191
		0.01	0.1130	13.18%	0.1353
		0.001	0.0944	10.58%	0.1086
90	10	0.1	0.0997	10.58%	0.1169
		0.01	0.0956	10.15%	0.1085
		0.001	0.0958	10.22%	0.1081

Berdasarkan Tabel 1 tersebut, akurasi LSTM tertinggi ada pada pembagian dataset 70:30 dengan learning rate 0.01. Dengan hasil MAE 0.0865, MAPE 9.49%, dan RMSE 0.1008. Berikut adalah grafik prediksinya.



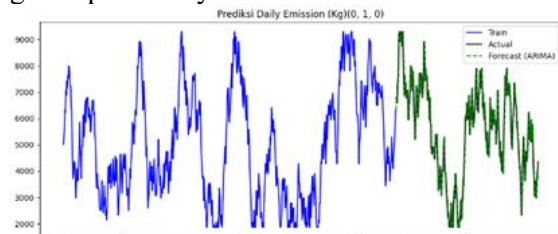
Gambar 8. Grafik Hasil Terbaik LSTM

4.3.2. ARIMA

Tabel 2. Hasil Evaluasi Prediksi ARIMA

Data Train (%)	Data Test (%)	MAE	MAPE	RMSE
70	30	0.0835	9.33%	0.0975
80	20	0.0940	10.37%	0.1086
90	10	0.0914	9.79%	0.1040

Berdasarkan Tabel 2 tersebut, akurasi ARIMA tertinggi ada pada pembagian dataset 70:30. Dengan hasil MAE 0.0835, MAPE 9.33%, dan RMSE 0.0975. Berikut adalah grafik prediksinya.



Gambar 9. Grafik Hasil Model Terbaik ARIMA

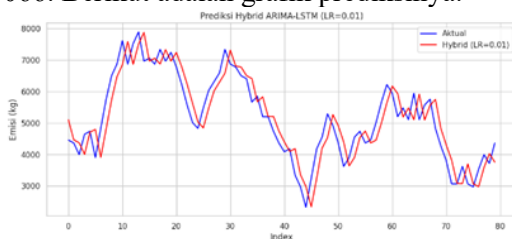
4.3.3. Hibrida LSTM dan ARIMA

Tabel 3. Hasil Evaluasi Prediksi Hibrida LSTM dan ARIMA

Data Train (%)	Data Test (%)	Lr	MAE	MAPE	RMSE
70	30	0.1	0.0951	10.39%	0.1081
		0.01	0.0941	10.20%	0.1066
		0.001	0.0946	10.26%	0.1071

80	20	0.1	0.1028	11.20%	0.1181
		0.01	0.1034	11.29%	0.1182
		0.001	0.1034	11.17%	0.1211
90	10	0.1	0.1103	11.31%	0.1233
		0.01	0.1114	11.44%	0.1244
		0.001	0.1119	11.51%	0.1237

Berdasarkan Tabel 3 tersebut, akurasi LSTM tertinggi ada pada pembagian dataset 70:30 dengan learning rate 0.01. Dengan hasil MAE 0.0941, MAPE 10.20%, dan RMSE 0.1066. Berikut adalah grafik prediksinya.



Gambar 10. Grafik Hasil Model Terbaik Hibrida LSTM dan ARIMA

5. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis penerapan model gabungan LSTM dan ARIMA dalam peramalan emisi karbon dioksida dari kendaraan bermotor, seraya menyoroti aspek ketepatan, kinerja, dan tingkat kesalahan model. Berdasarkan hasil eksperimen dan analisis data, beberapa kesimpulan utama berhasil dirumuskan, sebagaimana diuraikan berikut.

- a. Model Hibrida LSTM dan ARIMA menunjukkan kapabilitas yang memadai dalam memprediksi Emisi CO₂ Kendaraan Bermotor. Pada konfigurasi optimal (data latih 70%, *learning rate* 0.01), model ini mencatatkan nilai evaluasi MAE sebesar 0.0941, MAPE 10.20%, dan RMSE 0.1081. Hal ini merefleksikan keberhasilan model dalam mengintegrasikan kekuatan ARIMA untuk pola linier dan LSTM dalam mengolah pola *non-linier* dari residual,

sehingga memungkinkan adaptasi terhadap pola data Emisi CO₂ yang kompleks.

- b. Meskipun demikian, hasil pengujian mengindikasikan bahwa model hibrida belum mampu melampaui performa prediksi model ARIMA dan LSTM yang digunakan secara tunggal. Model ARIMA tercatat memberikan performa terbaik dengan MAE 0.0847, MAPE 9.21%, dan RMSE 0.0965, sementara model LSTM menghasilkan MAE 0.0912, MAPE 9.93%, dan RMSE 0.1032. Temuan ini menyiratkan bahwa optimalisasi pemanfaatan residual dari ARIMA sebagai masukan bagi LSTM masih perlu dipertimbangkan.
- c. Salah satu faktor pembatas yang mempengaruhi kinerja model hibrida adalah ketersediaan data yang terbatas serta minimnya variasi fitur dalam set data yang diaplikasikan. Dibandingkan dengan set data yang digunakan dalam beberapa publikasi ilmiah acuan, penelitian ini menggunakan jumlah observasi dan atribut yang lebih sedikit, sehingga dapat mengurangi kapasitas model dalam menangkap pola-pola kompleks dan jangka panjang secara maksimal.
- d. Meskipun belum menunjukkan performa superior secara keseluruhan, model hibrida ARIMA-LSTM tetap memiliki potensi signifikan sebagai pendekatan yang adaptif dan fleksibel, terutama untuk data deret waktu yang mengandung karakteristik linier dan non-linier secara simultan. Model ini diharapkan dapat berkinerja lebih optimal apabila diterapkan pada set data yang lebih besar, lebih kaya fitur, dan memiliki pola residual yang lebih jelas.
- e. Secara garis besar, penelitian ini mengindikasikan bahwa penerapan model Hibrida LSTM dan ARIMA dapat menjadi alternatif yang valid dalam prediksi emisi CO₂ kendaraan bermotor, dengan hasil evaluasi yang tergolong baik dan tingkat kesalahan yang masih berada dalam batas toleransi. Hasil ini juga membuka peluang untuk pengembangan lebih

lanjut, baik dari sisi arsitektur model, tahap *preprocessing* data, maupun strategi pelatihan, guna mencapai kapabilitas prediktif yang lebih tinggi.

UCAPAN TERIMA KASIH

Peneliti ingin menyampaikan rasa syukur yang sedalam-dalamnya kepada Tuhan Yang Maha Esa atas segala rahmat dan petunjuk-Nya, sehingga penulis berhasil menyelesaikan penelitian ini dengan baik. Penulis juga ingin menyampaikan terima kasih kepada kedua orang tua tercinta atas dukungan moril dan materiil yang sangat berarti, serta kepada semua pihak yang telah berpartisipasi dalam kelancaran penelitian ini

DAFTAR PUSTAKA

- [1] T. Wen, Y. Liu, Y. he Bai, and H. Liu, "Modeling and forecasting CO2 emissions in China and its regions using a novel ARIMA-LSTM model," *Heliyon*, vol. 9, no. 11, Nov. 2023.
- [2] IPCC, "2019 Refinement to The 2006 IPSS Guidelines for National Greenhouse Gas Inventories," 2019.
- [3] S. Anwar et al., "INDONESIA Third Biennial Update Report Under the United Nations Framework Convention on Climate Change REPUBLIC OF INDONESIA Coordinating Lead Authors Acknowledgement: Ministry of Environment and Forestry would like to thank to Ministry of Energy and Mineral Resources," 2021. [Online]. Available: <http://www.ditjenppi.menlhk.go.id>.
- [4] D. Castells-Quintana, E. Dienesch, and M. Krause, "Density, cities and air pollution: a global view," 2020. [Online]. Available: <https://ssrn.com/abstract=3713325>.
- [5] K. Mehmood et al., "The impact of COVID-19 pandemic on air pollution: a global research framework, challenges, and future perspectives," *Environmental Science and Pollution Research*, vol. 29, no. 35, pp. 52618–52634, Jul. 2022, doi: 10.1007/s11356-022-19484-5.
- [6] European Environment Agency, *The European Environment: State and outlook 2015*. Publications Office, 2015.
- [7] Rismawati Sihombing, "KAJIAN KEBUTUHAN JALUR HIJAU SEBAGAI PENYERAP EMISI KARBON DIOKSIDA KENDARAAN BERMOTOR DI SEPANJANG JALAN ARTERI MANOKWARI," 2012.
- [8] H. Sukarto, J. Teknik, H. Tower, L. Karawaci, Tangerang, and Banten, "PEMILIHAN MODEL TRANSPORTASI di DKI JAKARTA dengan ANALISIS KEBIJAKAN 'PROSES HIRARKI ANALITIK,'" vol. 3, Apr. 2006.
- [9] H. Yulianto, S. Iwan, and N. Daulay, "Analisis Efektivitas Metode Forecasting Terhadap Permintaan Produk Pt Arara Abadi Perawang Effectiveness Analysis Method Of Forecasting Product Demand Pt Arara Abadi Perawang," 2014.
- [10] T. Aljuneidi, S. Punia, A. Jebali, and K. Nikolopoulos, "Forecasting and planning for a critical infrastructure sector during a pandemic: Empirical evidence from a food supply chain," *Eur J Oper Res*, vol. 317, no. 3, pp. 936–952, 2024, doi: <https://doi.org/10.1016/j.ejor.2024.04.009>.
- [11] S. Rahayu and D. H. Nurdiansyaha, "Analisis Peramalan Penjualan Produk Kaos Sablon (Studi Kasus: Home Industry Alva Cloth)," *Jurnal Ilmiah Wahana Pendidikan*, vol. 8, no. 22, pp. 383–393, 2022, doi: 10.5281/zenodo.7349799.
- [12] A. A. T. Susilo, R. Kurniawan, and G. P. Julyansa, "PENERAPAN MODEL REGRESI LINEAR BERGANDA PADASISTEM PREDIKSI PRODUKSI HASIL PERTANIAN(PADI) DI KOTA LUBUKLINGGAU DENGAN BAHASA PEMROGRAMAN PYTHON," 2024.
- [13] Fungki Wahyu and Billy Hendrik, "Perbandingan Algoritma Time Series Dan Fuzzy Inference System Dalam Analisis Data Deret Waktu," *Jurnal Penelitian Teknologi Informasi dan Sains*, vol. 1, no. 3, pp. 16–24, Aug. 2023, doi: 10.54066/jptis.v1i3.711.
- [14] N. Mukaromah, "ANALISIS FORECASTING (PERAMALAN) PERMINTAANKARET PADA PT PERKEBUNAN NUSANTARA XII BANJARSARI JEMBER," 2019.
- [15] N. Zuchet and A. Orvieto, "Recurrent neural networks: vanishing and exploding gradients are not the end of the story," May 2024, [Online]. Available: <http://arxiv.org/abs/2405.21064>.
- [16] S. Siami-Namini, N. Tavakoli, and A. Siami Namin, "A Comparison of ARIMA and LSTM in Forecasting Time Series," in *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, Institute of Electrical and Electronics Engineers Inc., Jul. 2018, pp. 1394–1401. doi: 10.1109/ICMLA.2018.00227.
- [17] S. Wahyuningsih and dan Desi Yuniarti, "Penerapan Metode ARIMA Ensembel pada Peramalan (Studi Kasus: Inflasi di Indonesia)

- Application of ARIMA Ensemble Method in forecasting (Case Study: Inflation in Indonesia),” Jurnal EKSPONENSIAL, vol. 7, no. 1, 2016.
- [18] Z. Liu et al., “The Characteristics of ARMA (ARIMA) Model and Some Key Points to Be Noted in Application: A Case Study of Changtan Reservoir, Zhejiang Province, China,” Sustainability (Switzerland), vol. 16, no. 18, Sep. 2024, doi: 10.3390/su16187955.