

Peramalan pertumbuhan PDB Indonesia menggunakan hybrid ARIMAX-XGBoost dengan variabel ekspor-impor

Wanda Gustrifa*, Budi Nugroho, Achmad Junaedi

Program Studi Informatika, Fakultas Ilmu Komputer, UPN "Veteran" Jawa Timur
*Email: 21081010083@upnjatim.ac.id

Abstrak

Produk Domestik Bruto (PDB) merupakan indikator makroekonomi utama yang mencerminkan kinerja perekonomian suatu negara. Peramalan PDB yang akurat sangat diperlukan untuk perencanaan kebijakan strategis. Namun, karakteristik data PDB Indonesia sering kali menunjukkan pola volatilitas yang kompleks akibat pengaruh dinamika perdagangan global, yang sulit ditangkap sepenuhnya oleh model linear konvensional. Penelitian ini bertujuan untuk membandingkan kinerja peramalan model ekonometrika Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) dengan model Hybrid ARIMAX-XGBoost. Variabel eksogen yang digunakan difokuskan pada nilai Ekspor dan Impor untuk menangkap pengaruh aktivitas ekonomi eksternal. Data yang digunakan adalah PDB Atas Dasar Harga Konstan (ADHK) kuartalan yang telah melalui proses penyelarasan tahun dasar (backcasting) dan transformasi Year-on-Year (YoY). Metodologi hybrid dibangun dengan memodelkan komponen linear menggunakan ARIMAX, kemudian memodelkan residu (sisaan) model tersebut menggunakan algoritma Extreme Gradient Boosting (XGBoost) dengan parameter standar untuk menangkap pola non-linear. Hasil evaluasi pada data uji (testing set) dari periode 2020Q2 menunjukkan bahwa model Hybrid ARIMAX-XGBoost menghasilkan nilai Root Mean Squared Error (RMSE) dan Mean Absolute Error (MAE) yang lebih rendah dibandingkan model ARIMAX tunggal dan ARIMA standar. Penelitian ini menyimpulkan bahwa integrasi komponen non-linear efektif meningkatkan akurasi peramalan pertumbuhan ekonomi.

Kata Kunci: PDB; ARIMA; ARIMAX; ARIMAX-XGBoost; Ekspor-Import

Forecasting Indonesia's GDP growth using hybrid ARIMAX-XGBoost with export-import variables

Abstract

Gross Domestic Product (GDP) is the primary macroeconomic indicator reflecting a country's economic performance. Accurate GDP forecasting is essential for strategic policy planning. However, Indonesia's GDP data characteristics often exhibit complex volatility patterns due to global trade dynamics, which are difficult to fully capture by conventional linear models. This study aims to compare the forecasting performance of the Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) econometric model with the Hybrid ARIMAX-XGBoost model. Exogenous variables used are focused on Export and Import values to capture the influence of external economic activities. The data used is quarterly GDP at Constant Prices which has undergone through a base year alignment process (backcasting) and Year-on-Year (YoY) transformation. The hybrid methodology is constructed by modeling the linear component using ARIMAX, then modeling the residuals of that model using the standard Extreme Gradient Boosting (XGBoost) algorithm to capture non-linear patterns. Evaluation results on the testing set (from 2020Q2) show that the Hybrid ARIMAX-XGBoost model produces lower Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) values compared to the single ARIMAX and standar ARIMA model. This study concludes that integrating non-linear components effectively improves economic growth forecasting accuracy.

Keywords: GDP; ARIMA; ARIMAX; ARIMAX-XGBoost; Export-Import

1. Pendahuluan

Pertumbuhan ekonomi yang diukur melalui Produk Domestik Bruto (PDB) adalah tolok ukur fundamental evaluasi kesehatan ekonomi suatu negara (Xie et al., 2024). Akurasi dalam memproyeksikan angka pertumbuhan ini sangat krusial, baik bagi pemerintah dalam penyusunan Anggaran Pendapatan dan Belanja Negara (APBN) maupun bagi sektor swasta dalam pengambilan keputusan investasi (Kurniawan & Choir, 2024). PDB dipengaruhi oleh dinamika perdagangan

internasional, khususnya kinerja ekspor dan impor. Gejala pada sektor eksternal ini sering kali mentransmisikan volatilitas ke dalam perekonomian domestik.

Metode peramalan runtun waktu (time series) klasik seperti Autoregressive Integrated Moving Average (ARIMA) yang diperkenalkan oleh Box dan Jenkins telah lama menjadi standar dalam peramalan ekonomi (Jijo, 2025). Metode ini digemari karena ketegarannya dalam menangkap pola linear dan tren jangka pendek (Almaafi et al., 2023). Pengembangan model ini menjadi ARIMAX memungkinkan pelibatan variabel eksogen untuk meningkatkan daya jelas model terhadap fenomena ekonomi (Salsabila & Oktaviarina, 2024; Adhwaningrum et al., 2024). Namun, data ekonomi makro sering kali mengandung karakteristik non-linear dan kompleksitas yang tidak dapat dijelaskan sepenuhnya oleh asumsi linearitas (Agarwal, 2024). Pola non-linear ini sering kali muncul dalam bentuk residu (error) yang tidak acak pada model linear konvensional.

Untuk mengatasi keterbatasan tersebut, pendekatan hybrid yang menggabungkan model linear dan non-linear semakin banyak dikembangkan dalam literatur terkini (Odhiambo et al., 2024). Salah satu algoritma machine learning yang populer karena kinerjanya yang tinggi dan kecepatan komputasinya adalah Extreme Gradient Boosting (XGBoost) (Nichani et al., 2024). Konsep hybrid bekerja dengan prinsip dekomposisi di mana model linear menangkap tren utama, sedangkan model non-linear memodelkan residu yang tidak tertangkap (Zhang, 2003). Penelitian sebelumnya menunjukkan bahwa model hybrid memberikan hasil lebih baik daripada model tunggal dalam peramalan data finansial dan ekonomi (Kusumaningrum et al., 2023; Vonitsanos et al., 2024). Selain itu, teknik ini terbukti efektif dalam menangani data dengan tingkat noise yang tinggi.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk membandingkan akurasi peramalan antara model ARIMAX tunggal dengan model Hybrid ARIMAX-XGBoost. Penelitian ini secara spesifik menggunakan variabel ekspor dan impor sebagai prediktor eksogen.

2. Metode

Penelitian ini menggunakan pendekatan kuantitatif dengan data sekunder. Tahapan metodologi dijelaskan sebagai berikut:

2.1. Sumber data dan variabel

Data yang digunakan bersumber dari Badan Pusat Statistik (BPS) Indonesia dengan periode kuartalan mulai dari Kuartal I-1996 hingga Kuartal III-2025. Variabel yang digunakan meliputi:

- PDB ADHK (Y):** Produk Domestik Bruto Atas Dasar Harga Konstan.
- Ekspor (X1) dan Impor (X2):** Nilai ekspor dan impor barang dan jasa.

Pada data PDB ADHK Dilakukan proses *Backcasting* untuk menyelaraskan tahun dasar 1993, 2000 dan 2010 agar menjadi satu seri data yang konsisten. Penyamaan tahun dasar menggunakan rumus (Badan Pusat Statistik, 2010):

$$y_{t-1,j,d} = \frac{y_{t-1,j,d-1}}{y_{t,j,d-1}} \times y_{t,j,d}$$

Kemudian, seluruh data ditransformasikan ke dalam bentuk pertumbuhan *Year-on-Year* (YoY) menggunakan rumus:

$$\text{Growth} = \frac{\text{value}_t - \text{value}_{t-4}}{\text{value}_{t-4}} \times 100$$

Transformasi ini bertujuan untuk menstabilkan varians data (stasioneritas) dan menghilangkan efek musiman tahunan.

2.2. Desain pemodelan

Data dibagi menjadi dua bagian, yaitu data latih (*training set*) untuk pembentukan model dan data uji (*testing set*) untuk evaluasi kinerja peramalan. Pada penelitian ini data test dilakukan pada periode Q2 2020 – Q3 2025.

a. Tahap 1: Pemodelan ARIMAX

Model ARIMAX dibangun untuk menangkap pola linear. Variabel Ekspor dan Impor dimasukkan sebagai variabel eksogen. Identifikasi ordo p,d,q dilakukan melalui plot *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF). Pemilihan model terbaik didasarkan pada nilai *Akaike Information Criterion* (AIC) terkecil. Persamaan model ARIMAX adalah (Vonitsanos et al., 2024):

$$\hat{Y}_{ARIMAX} = c + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \beta X_t + \varepsilon_t$$

b. Tahap 2: Pemodelan Hybrid

Residu (e_t) yang dihasilkan dari model ARIMAX dihitung selisihnya terhadap data aktual ($e_t = Y_{Aktual} - \hat{Y}_{ARIMAX}$) (Zhang, 2003). Residu ini diasumsikan mengandung informasi non-linear. Algoritma XGBoost dengan parameter *baseline* digunakan untuk memprediksi nilai residu tersebut. Prediksi akhir diperoleh dengan menjumlahkan prediksi ARIMAX dan prediksi residu dari XGBoost:

$$\hat{Y}_{Hybrid} = \hat{Y}_{ARIMAX} + \hat{e}_{XGBoost}$$

2.3. Evaluasi Kinerja

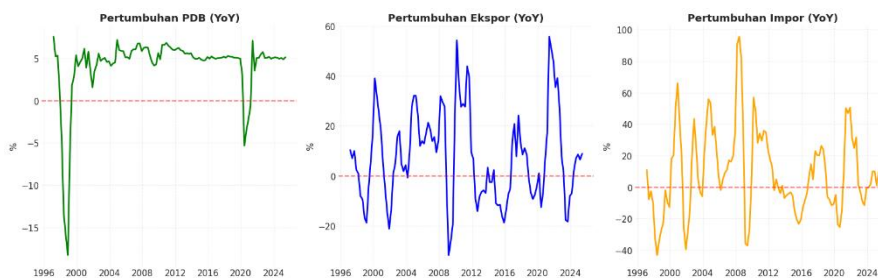
Akurasi peramalan dievaluasi menggunakan *Root Mean Squared Error* (RMSE) dan *Mean Absolute Error* (MAE) pada data *testing*. Semakin kecil nilai metrik, semakin baik kinerja model.

3. Hasil dan Pembahasan

Bagian ini memaparkan hasil analisis data, mulai dari karakteristik data hingga perbandingan kinerja model.

3.1. Karakteristik Data dan Uji Stasioneritas

Analisis diawali dengan mengamati perilaku data runtun waktu (*time series*) untuk memahami karakteristik dasar pertumbuhan ekonomi dan perdagangan internasional Indonesia. Visualisasi tren pertumbuhan *Year-on-Year* (YoY) disajikan pada **Gambar 1**.



Gambar 1. Plot Time Series Pertumbuhan PDB, Ekspor, dan Impor

Pada **Gambar 1** menunjukkan adanya fluktuasi pada pergerakan pertumbuhan PDB, ekspor dan impor. Fluktuasi tersebut mengindikasikan ada pola non-linear yang dimana pola tersebut tidak dapat ditangkap oleh model ARIMAX sehingga penerapan model hybrid pada data akan lebih efektif dan akurat.

Sebelum pemodelan, dilakukan uji stasioneritas menggunakan *Augmented Dickey-Fuller* (ADF) (Gujarati & Damodar, 2004). Uji stasioner ini perlu dilakukan dalam analisis data deret waktu untuk meminimalisir ketidakakuratan model. Apabila data tidak stasioner maka perlu dilakukan differencing (Box & Cox, 1964). Hasil uji menunjukkan bahwa setelah transformasi YoY, nilai probabilitas (p-value) untuk ketiga variabel bernilai kurang dari 0.05.

Tabel 1. Hasil Uji Stasioneritas (ADF)

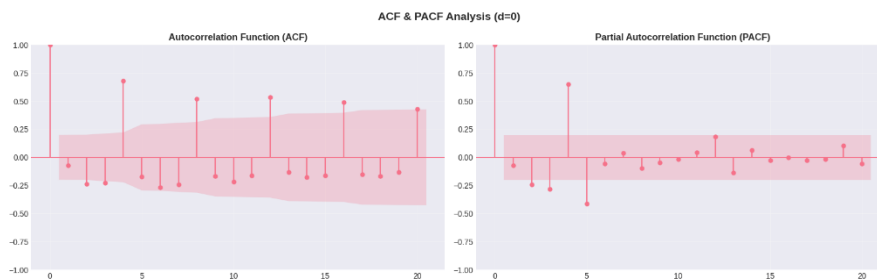
Variabel	Statistik ADF	P-Value	Kesimpulan
PDB (YoY)	-3.3379	0.0133	Stasioner
Ekspor (YoY)	-3.1372	0.0239	Stasioner
Impor (YoY)	-4.447	0.0002	Stasioner

Sumber: Hasil Olah Data (2026)

Hasil pada **Tabel 1** mengonfirmasi bahwa data sudah stasioner dan memenuhi syarat untuk pemodelan ARIMAX tanpa perlu *differencing* lebih lanjut.

3.2. Pemodelan ARIMAX

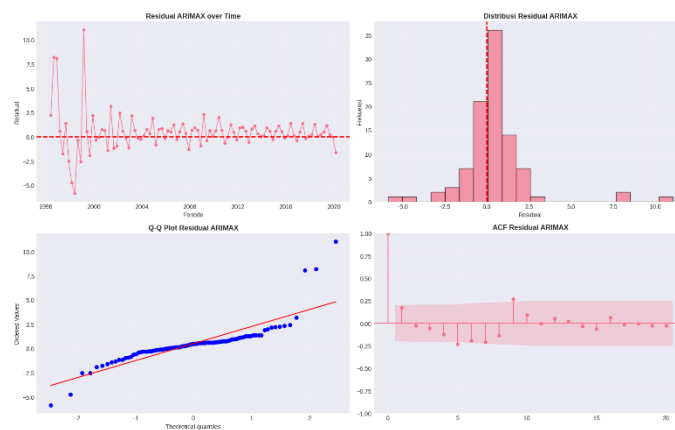
Pemodelan tahap pertama menggunakan ARIMAX bertujuan untuk menangkap hubungan linear antara variabel dependen (PDB) dengan variabel independen (Ekspor dan Impor). Setelah data dipastikan stasioner, selanjutnya yaitu estimasi parameter menggunakan ACF dan PACF.



Gambar 2. Plot ACF dan PACF

Pada **Gambar 2** diperoleh $p = [0,5]$ dan $q = [0,4]$. Model-model yang terbentuk berdasarkan ACF dan PACF dilakukan uji signifikansi parameter. Namun, model-model yang terbentuk tidak ada yang memiliki semua parameter signifikan maka memilih model yang menggunakan *Akaike Information Criterion* (AIC) terkecil (Ugoh et al., 2021). Berdasarkan kriteria AIC terkecil, diperoleh model (5, 0, 3) dengan AIC sebesar 375.46 yang akan digunakan pada pemodelan ARIMAX.

Selanjutnya, dilakukan uji diagnostik terhadap residu model ARIMAX. Hasil pengujian diagnostik seperti pada **Gambar 3**.



Gambar 3. Uji Diagnostik

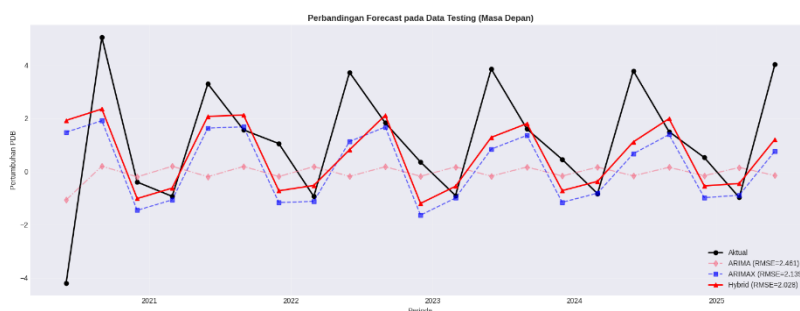
Pada **Gambar 3** hasil analisis diagnostik menunjukkan bahwa residu masih memiliki fluktuasi yang cukup tinggi pada periode-periode tertentu, yang mengindikasikan adanya pola yang belum tertangkap sempurna (*white noise* belum tercapai sepenuhnya). Hal ini menjadi dasar justifikasi penerapan model *hybrid*.

3.3. Pemodelan Hybrid ARIMAX-XGBoost

Pada tahap ini, residu dari model ARIMAX diprediksi menggunakan XGBoost. Fitur yang digunakan dalam XGBoost adalah *lag* dari residu itu sendiri. XGBoost mampu menangkap pola non-linear pada residu, terutama pada saat terjadi perubahan arah tren yang drastis. Setelah residu diperbaiki oleh XGBoost, dilakukan pemodelan hybrid dengan menambahkan perbaikan residu terhadap arimax forecasting.

3.4. Perbandingan Kinerja Peramalan

Setelah dilakukan pemodelan ARIMAX dan Hybrid, diterapkan model ARIMA standar sebagai baseline pembanding untuk melihat pengaruh ekspor dan impor terhadap akurasi dalam peramalan pertumbuhan PDB. Hasil peramalan ketiga model dibandingkan dengan data aktual pada data latih. Visualisasi perbandingan dapat dilihat pada **Gambar 4**.



Gambar 4. Grafik perbandingan data aktual vs ARIMA vs ARIMAX vs Hybrid

Secara visual, garis prediksi model *Hybrid* (warna merah) terlihat lebih mendekati data aktual dibandingkan model ARIMAX (warna biru) dan ARIMA (warna oranye), terutama dalam merespons pemulihan pasca guncangan ekonomi. Untuk validasi kuantitatif, **Tabel 2** menyajikan nilai error model.

Tabel 2. Evaluasi Kinerja Model pada Data Latih

Model	RMSE	MAE (%)
Hybrid ARIMAX-XGBoost	2.028	1.467
ARIMAX	2.139	1.524
ARIMA	2.461	2.002

Sumber: Hasil Olah Data (2026)

Berdasarkan **Tabel 3**, model *Hybrid* ARIMAX-XGBoost menghasilkan nilai RMSE sebesar 2.028 dan MAE sebesar 1.467%. Nilai ini lebih rendah dibandingkan model ARIMAX tunggal dan ARIMA standar yang masing-masing memiliki RMSE 2.139; MAE 1.524% dan RMSE 2.461; MAE 2.002%. Penurunan nilai error ini menunjukkan bahwa penambahan komponen non-linear melalui XGBoost serta penggunaan ekspor-impor sebagai variabel eksogen berhasil memperbaiki akurasi peramalan. Mekanisme koreksi residu yang dilakukan oleh XGBoost mampu meminimalisir kesalahan sistematis yang dihasilkan oleh model linear.

4. Kesimpulan

Variabel Ekspor dan Impor dapat meningkatkan akurasi terhadap peramalan pertumbuhan PDB Indonesia.

Model linear ARIMAX memiliki keterbatasan dalam menangkap volatilitas penuh dari data pertumbuhan ekonomi, yang tercermin dari residu yang masih mengandung pola.

Model *Hybrid* ARIMAX-XGBoost terbukti lebih unggul dibandingkan model ARIMAX tunggal dan ARIMA standar. Pendekatan *hybrid* mampu menurunkan nilai RMSE dan MAE, yang mengindikasikan peningkatan presisi peramalan.

Implikasi dari penelitian ini adalah bahwa penggunaan metode gabungan (statistik dan *machine learning*) sangat disarankan dalam proyeksi ekonomi makro, bahkan tanpa perlu melakukan optimasi hyperparameter yang rumit, model *baseline* XGBoost sudah mampu memberikan kontribusi positif terhadap akurasi model linear.

5. Ucapan terimakasih

Penulis mengucapkan terima kasih kepada Badan Pusat Statistik (BPS) atas penyediaan data publik yang berkualitas.

Daftar Pustaka

- Adhwaningrum, A. S., Amri, I. F., Dwi Saputri, A., Diani, N. L., Pratama, R. F., & Al Haris, M. (2024). Perbandingan Model Arima Dan Arimax Untuk Peramalan Temperatur Di Kota Semarang. *Amalgamasi: Journal of Mathematics and Applications*, 3(2), 54–64. <https://doi.org/10.55098/amalgamasi.v3.i2.pp54-64>
- Agarwal, P. (2024). *Enhancing Financial Forecasting in ERP Systems using XGBoost: A Robust Sales Prediction Model* (Vol. 2024, Issue Icetsbp). Atlantis Press International BV. https://doi.org/10.2991/978-94-6463-544-7_26
- Almaafi, A., Bajaba, S., & Alnori, F. (2023). Stock price prediction using ARIMA versus XGBoost models: the case of the largest telecommunication company in the Middle East. *International Journal of Information Technology*, 15(4), 1813–1818. <https://doi.org/10.1007/s41870-023-01260-4>
- Badan Pusat Statistik. (2010). *Katalog BPS: 9301004*.
- Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. In *Journal of the Royal Statistical Society Series B: Statistical Methodology* (Vol. 26, Issue 2, pp. 211–243). <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>
- Gujarati, & Damodar, N. (2004). *Basic Econometrics* (4th ed.). McGraw-Hill Inc.
- Jijo, J. J. (2025). Modelling and Forecasting the GDP of G7 Countries Using Arima Model. *Asian Journal of Economics, Business and Accounting*, 25(3), 295–315. <https://doi.org/10.9734/ajebe/2025/v25i31711>
- Kurniawan, T. A., & Choir, A. S. (2024). Nowcasting Produk Domestik Bruto Atas Dasar Harga Konstan Triwulanan Indonesia. *Seminar Nasional Official Statistics*, 2024(1), 691–700. <https://doi.org/10.34123/semnasoffstat.v2024i1.2185>
- Kusumaningrum, N., Purnamasari, I., & Siringoringo, M. (2023). Peramalan Menggunakan Model Hybrid ARIMAX-NN untuk Total Transaksi Pembayaran Nontunai. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 5(1), 1–14. <https://doi.org/10.35580/variansiunm57>
- Nichani, R., Gasmi, L., Laiche, N., & Kabou, S. (2024). Optimizing financial time series predictions with hybrid ARIMA, LSTM, and XGBoost Models. *Studies in Engineering and Exact Sciences*, 5(2), e11188. <https://doi.org/10.54021/seesv5n2-582>
- Odhiambo, S. O., Cornelious, N. O., & Waititu, H. (2024). Developing a Hybrid ARIMA-XGBOOST Model for Analysing Mobile Money Transaction Data in Kenya. *Asian Journal of Probability and Statistics*, 26(10), 108–126. <https://doi.org/10.9734/ajpas/2024/v26i10662>
- Salsabila, N., & Oktaviarina, A. (2024). Peramalan PDRB PERAMALAN PDRB DI JAWA TIMUR MENGGUNAKAN MODEL ARIMAX DENGAN VARIABEL EKSOGEN EKSPOR-IMPOR. *MATHunesa: Jurnal Ilmiah Matematika*, 12(1), 208–218. <https://doi.org/10.26740/mathunesa.v12n1.p208-218>
- Ugoh, C., Uzuke, C., & Ugoh, D. (2021). Application of ARIMAX Model on Forecasting Nigeria's GDP. *American Journal of Theoretical and Applied Statistics*, 10, 216. <https://doi.org/10.11648/j.ajtas.20211005.12>
- Vonitsanos, G., Kanavos, A., & Mylonas, P. (2024). *A Context-Enriched Hybrid ARIMAX – Deep Learning Framework for Robust Cryptocurrency Price Forecasting*.
- Xie, H., Xu, X., Yan, F., Qian, X., & Yang, Y. (2024). *Deep Learning for Multi-Country GDP Prediction: A Study of Model Performance and Data Impact*. 1–13.

<http://arxiv.org/abs/2409.02551>

Zhang, P. (2003). Zhang, G.P.: Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Neurocomputing* 50, 159-175. *Neurocomputing*, 50, 159–175.
[https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)