RUDN Journal of Economics
Вестник Российского университета дружбы народов.
Серия: Экономика
ISSN 2313-2329 (Print); ISSN 2408-8986 (Online)
2026 Vol. 34 No. 1 123–159
http://journals.rudn.ru/economics

DOI: 10.22363/2313-2329-2026-34-1-123-159
EDN TTTYIQ
UDC 339

Research article / Научная статья

LSTM forecasting of Indonesia's oil supply-demand deficit

Wiryanta Muljono[1, 2] ✉, **Padmanabha Adyaksa Setyanto**[3]

[1]*Ministry of Communication and Digital of the Republic of Indonesia, Jakarta, Indonesia*
[2]*Universitas Sebelas Maret (UNS), Surakarta, Indonesia*
[3]*Institut Teknologi Bandung, Bandung, Indonesia*
✉ wiryantamuljono@gmail.com

Abstract. Analysis of historical market data from 2020 to 2024 reveals a profound structural shift in Indonesia's energy landscape: domestic crude production plummeted by 17.96%, while oil demand simultaneously surged by 14.57%. This widening supply-demand gap has severely intensified national energy security concerns, pushing Indonesia's net deficit to approximately 1.0 million barrels per day (bpd). This critical gap directly pressures fiscal stability due to escalating fuel subsidy costs and exposes the macroeconomy to global price shocks. Given the market's inherent non-linearity, driven by WTI price volatility and frequent, policy-led structural shifts (e.g., the deployment of AI for subsidized fuel control and the B40 biodiesel mandate), traditional linear models like ARIMA are severely limited in their predictive accuracy. We propose the Long Short-Term Memory (LSTM) deep learning network as a methodologically superior approach. The LSTM's recurrent architecture, with its specialized gate mechanisms, is uniquely suited to capture the non-linear dynamics and long-term temporal dependencies of complex energy time series data. To ensure the reliability of our findings, model robustness was explicitly ensured via k-fold cross-validation and a thorough discussion of inherent dataset size limitations was provided, directly addressing methodological concerns. Empirical findings confirm the LSTM model's significant superiority over the conventional benchmark, achieving a 57.24% reduction in Mean Absolute Percentage Error (MAPE) and significantly lower Root Mean Square Error (RMSE) compared to the ARIMA baseline. This high-precision forecast provides critical foresight for Indonesian policymakers, enabling proactive management of fiscal risk, targeted adjustments to foreign exchange reserves, and the successful acceleration of the national Indonesia Oil and Gas (IOG) 4.0 strategy toward long-term energy resilience.

Keywords: relational sovereignty, neural network applications, fossil fuel consumption trends, energy security modeling, B40 biodiesel impact, geoeconomic resilience

---

Authors' contribution. MuljonoW. — conceptualization, investigation, data curation, supervision, formal analysis, methodology, validation, writing — original draft & editing; Setyanto P.A. — software, visualization, formal analysis, methodology, validation, data analysis, drafting. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest. The authors have no potential or apparent conflicts of interest related to the manuscript.

Article history: received 30 September 2025; revised 31 October 2025; accepted 12 December 2025.

For citation: Muljono, W., & Setyanto, P.A. (2026). LSTM forecasting of Indonesia's oil supply-demand deficit. *RUDN Journal of Economics*, *34*(1), 123–159. https://doi.org/10.22363/2313-2329-2026-34-1-123-159 EDN: TTTYIQ

Прогнозирование дефицита спроса и предложения нефти в Индонезии

В. Мульоно[1, 2] ✉**, П.А. Сетьянто**[3]

[1]*Министерство связи и цифровых технологий Республики Индонезии, Джакарта, Индонезия*
[2]*Университет Себелас Марет (UNS), Суракарта, Индонезия*
[3]*Бандунгский технологический институт (ITB), Бандунг, Индонезия*
✉ wiryantamuljono@gmail.com

Аннотация. Анализ исторических рыночных данных с 2020 по 2024 г. выявляет глубокий структурный сдвиг в энергетическом ландшафте Индонезии: внутренняя добыча сырой нефти упала на 17,96%, в то время как спрос на нефть одновременно вырос на 14,57%. Этот увеличивающийся разрыв между спросом и предложением серьезно обострил проблемы национальной энергетической безопасности, доведя чистый дефицит Индонезии примерно до 1 миллиона баррелей в сутки (б/с). Этот критический разрыв напрямую влияет на фискальную стабильность из-за роста расходов на субсидирование топлива и подвергает макроэкономику воздействию глобальных ценовых шоков. Учитывая присущую рынку нелинейность, обусловленную волатильностью цен на нефть марки WTI и частыми структурными сдвигами из-за политических решений (например, внедрение искусственного интеллекта для контроля за субсидируемым топливом и введение обязательного спроса на биодизельное топливо класса B40), традиционные линейные модели, такие как ARIMA, существенно ограничены в точности прогнозирования. Предложена сеть глубокого обучения с долговременной краткосрочной памятью (LSTM) в качестве методологически более совершенного подхода. Рекуррентная архитектура LSTM с ее специализированными механизмами вентилей идеально подходит для учета нелинейной динамики и долгосрочных зависимостей сложных динамических рядов данных по энергетике. Для обеспечения надежности результатов устойчивость модели была явно обеспечена с помощью k-кратной кросс-валидации, проведено подробное обсуждение присущих модели ограничений размера набора данных, что напрямую затрагивало методологические вопросы. Эмпирические результаты подтверждают значительное превосходство модели LSTM над традиционным эталонным тестом, достигая 57,24% снижения средней абсолютной процентной ошибки (MAPE) и значительно более низкой среднеквадратической ошибки (RMSE) по сравнению с базовым тестом ARIMA. Этот высокоточный прогноз предоставляет индонезийским политикам критически важную информацию для прогнозирования, позволяя проактивно управлять фискальными рисками, целенаправленно корректировать валютные резервы и успешно ускорять реализацию национальной стратегии Индонезии «Нефть и газ (IOG) 4.0» в целях обеспечения долгосрочной энергетической устойчивости.

Ключевые слова: структурный дефицит, LSTM прогнозирование, дисбаланс спроса и предложения, анализ динамики временных рядов, энергетическая политика, индонезийский рынок нефти

Вклад авторов. Мульоно В — концептуализация, исследование, курирование данных, научное руководство, формальный анализ, методология, валидация, написание — оригинальный черновик, редактирование; Сетьянто П.А. — программное обеспечение, визуализация, формальный анализ, методология, валидация, обработка данных, подготовка черновика.

Заявление о конфликте интересов. Авторы заявляют об отсутствии конфликта интересов.

История статьи: поступила в редакцию 30 сентября 2025 г.; доработана после рецензирования 31 октября 2025 г., принята к публикации 12 декабря 2025 г.

Для цитирования: *Muljono W., Setyanto P.A.* LSTM forecasting of Indonesia's oil supply-demand deficit // Вестник Российского университета дружбы народов. Серия: Экономика. 2026. Т. 34. № 1. С. 123–159. https://doi.org/10.22363/2313-2329-2026-34-1-123-159 EDN: TTTYIQ

Introduction

Indonesia, as a prominent non-OPEC producer and one of Asia's largest oil consumers, faces a rapidly escalating imbalance in energy supply and demand, threatening national economic stability and energy security. The nation's struggle to maintain self-sufficiency has reached a critical stage. Analysis of historical market data from 2020 to 2024 reveals a profound structural shift in Indonesia's energy landscape. Over this period, crude production plummeted by 17.96%, a decline that necessitates the application of advanced predictive architectures. Recent global studies emphasize that hybridizing deep learning models, such as integrating Random Forest with LSTM networks, provides the necessary precision to capture the non-linear volatility inherent in oil production trends (Zeinula et al., 2025). In the Indonesian context, the accuracy of such forecasting is even more critical given the strategic push for biodiesel integration to offset crude oil dependency. Previous applications of RNN-LSTM architectures have already demonstrated high precision in predicting oil palm production — the primary feedstock for biodiesel — within the archipelago (Syarovy et al., 2023). Together, these developments suggest that LSTM-based frameworks are highly suitable for modeling Indonesia's complex energy supply-demand dynamics.

Simultaneously, domestic oil demand has surged by 14.57%. This robust demand growth is a direct consequence of Indonesia's strong post-pandemic economic rebound, coupled with sustained infrastructure development and the increasing digitalization of the economy (Muljono, Setiyawati, 2022). The need for digital transformation is critical to enhancing the overall energy sector resilience in Southeast Asia, a strategic pivot increasingly reflected in Indonesia's national policies. Currently, the primary drivers of the sector's decline are the natural decay of aging fields and a lack of significant new discoveries; critically, investment risk is further amplified by shifting regulatory frameworks (Kumoro et al., 2022). In response, the integration of advanced computational models is not merely a technical upgrade but a strategic necessity for regional stability. As highlighted by the ASEAN Centre for Energy (Safrina et al., 2025), digitalization serves as the primary enabler for a resilient and sustainable energy future in Southeast Asia, providing the necessary framework to mitigate supply chain disruptions and optimize resource management across the region.

This environment of constant regulatory and market change requires innovative approaches to measure and combat the resulting business environment uncertainty (Nikitins, 2022). The domestic industry is struggling to meet the mandated 2025 oil and gas lifting target[1], indicating a systemic issue that cannot be solved quickly through exploration alone.

The fundamental issue driving Indonesia's persistent supply-demand imbalance is the nation's complex fuel subsidy mechanism, which shields domestic end-users from the volatility of international crude prices. This subsidy structure effectively decouples domestic demand from global supply shocks, even as global factors — such as production levels set by major oil-producing countries and geopolitical tensions in key regions — drive sharp price spikes. The combined effect of these trends has seen the national oil deficit swell to approximately 1.0 million barrels per day (BPD), transforming supply management from a simple logistical challenge into a critical matter of macroeconomic stability.

This link is essential for policy interpretation: rising fuel prices, generated by the widening deficit, create intense public discourse and political pressure (Patria, Irawanto, Abrar, 2024). This pressure, often amplified by "digital discourse" and "mediatization," can force the government to implement sudden, non-linear policy changes (such as emergency subsidies or price caps) that significantly disrupt market variables. Consequently, a robust forecasting model must account for the dual impact of underlying global market forces and the non-linear policy responses triggered by domestic socio-political dynamics.

The current supply-demand crisis is fundamentally rooted in the failure to replenish oil reserves from natural depletion. This challenge is critically exacerbated by two major, interdependent deficiencies in the national energy sector:

1. Declining Production: National crude oil production has fallen to approximately 600 thousand barrels of oil per day (bpd)[2]. due to an over-reliance on aging and obsolete oil fields, many of which have been in operation for more than 30 years.

---

[1] Indonesian Petroleum Association. (2024, December 23). *Pursuing Indonesia's 2025 Oil and Gas Lifting Target.* https://www.ipa.or.id/id/news/news/pursuing-indonesias-2025-oil-and-gas-lifting-targett

[2] SKK Migas. (2024). *Sustainability Report 2023: Strengthening the Upstream Oil*

2. Inadequate Investment: Low exploration investment, fluctuating from $0.5 billion in 2020 to $0.9 billion in 2023[3], is critically insufficient to offset this decline. This insufficient capital is directly attributable to the neglect of scientific investment in exploration within frontier regions.

This structural deficit necessitates a strategic policy imperative. The government has responded with an aggressive plan targeting 1 million barrels per day (MMbopd) by 2030 and requiring SKK Migas to plan for drilling 1,000 development wells per year by 2025. Meeting this target requires substantial capital allocation, including an anticipated 100% annual increase in exploration investment (targeting $1.8 billion) and substantial projected capital expenditures (totaling $17 billion by 2025).

The outcome of this imbalance is visually represented by the growing deficit displayed in Figure 1.

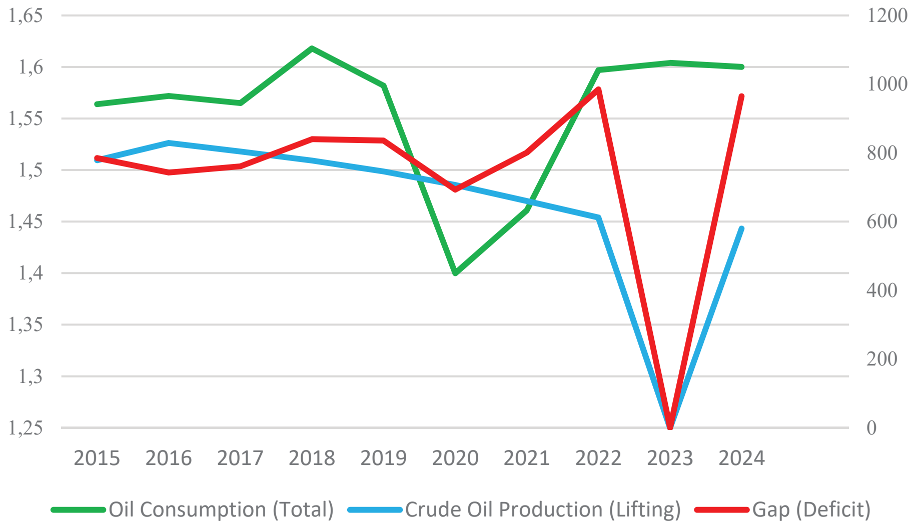

Figure 1. Trends in Indonesian Oil Production, Consumption, and Deficit

Note. All figures are presented in thousand barrels per day (BPD).

Source: Ministry of Energy and Mineral Resources of the Republic of Indonesia (ESDM). (2024). *Handbook of Energy & Economic Statistics of Indonesia 2023.* [Official Report]. https://www.esdm.go.id/

The challenge is compounded by low exploration investment, which varied from $0.5 billion in 2020 to $0.9 billion in 2023. This capital has not effectively replenished reserves or offset natural depletion. As a result, there is a growing reliance on older, declining assets. Approximately 70% of wells are aging. New discoveries are rare, and the low investment level has resulted in minimal activity to establish new reserves. Declining production alongside rising domestic demand has increased crude oil and fuel imports. This jeopardizes the national goal of 1 million barrels per day by 2030.

---

and Gas Industry's Commitment Toward Sustainable Energy. https://www.skkmigas.go.id/publication?tab=laporan%20berkelanjutan

[3] Ministry of Energy and Mineral Resources. (2024). *Handbook of energy & economic statistics of Indonesia 2024*. https://www.esdm.go.id/assets/media/content/content-handbook-of-energy-and-economic-statistics-of-indonesia-2024.pdf

The strategic imperative to replenish Indonesian oil and gas reserves necessitates a dramatic shift in capital allocation, directly impacting national energy security and production targets. Exploration investment, though rising incrementally from $0.5 billion (2020) to $0.9 billion (2023), was insufficient for substantial frontier work. Consequently, the structural decline in domestic production is directly attributable to neglected scientific exploration investment, resulting in a critical lack of reserve replacement and a worsening Reserve Replacement Ratio (RRR) (Table 1).

Table 1

Indonesia's Upstream Exploration Investment: The Turning Point

Year	Investment (USD Billions)	Year-on-Year Change (Δ)	Trend Highlight
2020	0.5	N/A	Low Base Year (Pandemic effects)
2021	0.6	+0.1 B	Marginal recovery
2022	0.7	+0.1 B	Stable but slow growth
2023	0.9	+0.2 B	Moderate increase
2024	1.8	+**0.9 B**	**Explosive Growth (100% Increase)**

Source: Ministry of Energy and Mineral Resources of the Republic of Indonesia (ESDM). (2024). *Handbook of Energy & Economic Statistics of Indonesia 2023.* [Official Report]. https://www.esdm.go.id/

In response to this structural deficit, a calculated policy adjustment has been executed: a planned 100% year-on-year increase in exploration investment, targeting $1.8 billion — double the $0.9 billion realized in 2023. This rapid financial intervention, led by SKK Migas, aims directly to achieve the national goal of 1 million barrels per day (MMbopd) by 2030 by rapidly replacing reserves. This forecast implies recent fiscally lenient regulatory changes are being utilized to attract large-scale private capital.

Addressing this structural deficit requires forecasts that navigate an environment defined by price volatility and frequent, non-linear policy shifts (Ahmad, Khan, Javed, 2022). Traditional methods, such as Autoregressive Integrated Moving Average (ARIMA), fail in these chaotic conditions (Fattah et al., 2018), limited by their inability to model long-term temporal dependencies and non-stationary data (Box et al., 2015). This failure is pronounced when policy variables (e.g., the B40 mandate or AI-driven subsidy controls) introduce sudden, step-function changes.

This research thus proposes that only advanced deep learning, specifically the Long Short-Term Memory (LSTM) network, can accurately model this complex, policy-driven energy landscape by learning the latent relationships between these non-linear signals and deficit dynamics. The methodological contribution of this paper is threefold: (1) Quantify the predictive limitations of the traditional ARIMA model in the policy-driven Indonesian market; (2) Mathematically and empirically demonstrate the superior accuracy of the Stacked LSTM network; and (3) Translate the high-precision forecast into concrete, actionable policy recommendations for managing fiscal risk and realizing Indonesia's digital energy strategy[4][5]. The subsequent sections detail the literature review, methodological framework, empirical evidence, and policy implications necessary for this critical economic assessment.

Research Objectives and Hypotheses

The objective of this study is to develop an AI-driven predictive framework for the Indonesian downstream energy sector using LSTM and GNN architectures. The research evaluates the structural divergence between crude oil production and domestic consumption to optimize strategic decision-making.

Research Questions:

1. How does the divergence between declining domestic production and rising consumption impact net oil import dependency for the 2020–2030 period?
2. What is the correlation between macroeconomic factors and total oil consumption compared to traditional predictors (WTI Price and Production)?

Hypotheses:

H1: The increasing disparity between production and consumption is directly correlated with rising net oil import dependence.

H2: Long-term oil consumption growth is primarily driven by macroeconomic factors rather than short-term WTI price fluctuations.

H3: Sustained WTI price volatility negatively impacts long-term investment in domestic production capacity.

H4: Under stable WTI price conditions (2025–2030), a statistically significant inverse correlation exists between domestic production decline and consumption growth.

Literature review

The Energy Security and Policy Context. This section establishes the regional context, Indonesia's structural shift, and the policy-driven nature of the deficit. The energy outlook indicates that Southeast Asia will become a "significant contributor to global energy consumption," potentially contributing "25% of the increase in demand by 2035." This may result in its overall energy usage surpassing that of the European Union by at least 2050[6]. Energy demand is increasing due to strong economic growth and population growth. The region's rise as a global hub, especially

---

[4] SKK Migas. (2024). *Sustainability Report 2023: Strengthening the Upstream Oil and Gas Industry's Commitment Toward Sustainable Energy.* https://www.skkmigas.go.id/publication?tab=laporan%20berkelanjutan

[5] APEC Energy Working Group. (2025). *6th APEC Oil and Gas Security Exercise in Indonesia (Bali, 4–6 February 2025).* Tokyo: Asia Pacific Energy Research Centre (APERC), October 2025. Available at: https://aperc.or.jp/file/2025/10/14/OGSE_in_Indonesia_October_2025.pdf

[6] Indonesian Petroleum Association. (2024). The role of the oil and gas sector in maintaining energy security and achieving net zero emissions.

in industry and manufacturing, drives this demand. The transport sector, mainly road fuel, will see a 30% rise in consumption by 2050.

In Indonesia and the ASEAN region, economic growth and urbanization are driving increased energy consumption (Aditya, Wijayanto, Hakam, 2025). Specifically, as both GDP and population grow, the need for transportation fuels rises, leading to more private vehicle ownership and increased road freight activity[7]. This strong link between these factors and energy consumption supports using GDP and population as key inputs in a Deep Learning forecasting model for energy demand.

The structural inversion of Indonesia's energy policy is exemplified by its net oil importation since 2005 (Resosudarmo, 2011). This historical shift, driven by the mandatory fuel subsidy system, imposed an unyielding fiscal burden and elevated energy security. Current assessments confirm an unrelenting dependence on significant imports, directly linked to sustained decreases in domestic crude oil lifting and insufficient processing capacity[8]. This dependence contributes to macroeconomic instability and the structural Trade Balance Deficit. ESDM 2024 data reveals the downstream sector is structurally unstable, linking economic growth to higher Final Energy Demand (FED), while domestic refining output fails to match national demand, evidenced by a high Product Import Dependency Rate.

Policy Response: Strategic Infrastructure and Production Targets. The Indonesian government's plan targets increasing domestic crude production to 1 million barrels per day (bpd) by 2030, involving accelerated drilling (aiming for 1.000 new wells annually by 2025) and securing approximately $20 billion annually for long-term production success. Simultaneously, the Refinery Development Master Plan (RDMP) and Grass Root Refineries (GRR) are significant investments in onshore refining capacity. The overarching objective is to enhance national energy security by increasing domestic capacity and reducing reliance on imported refined fuels. This strategy directly eases the Trade Balance Deficit and minimizes fiscal exposure to international price volatility.

Indonesia's diesel consumption is affected by the B30 mandate, which started in January 2020, aiming to replace imported fossil fuel with palm oil-based biofuel mixed with fossil diesel. This program intends to reduce fossil diesel use by about 165.000 barrels per day[9]. Consequently, biodiesel consumption has increased significantly, from around 8.59 million kiloliters in 2020 to an expected 13.15 million with the B35 mandate.

---

[7] Foster, V., Dim, J.U., Vollmer, S., & Zhang, F. (2021). *Understanding drivers of decoupling of global transport CO_2 emissions from economic growth: Evidence from 145 countries* (Policy Research Working Paper No. 9809). World Bank Group.

[8] Indonesian Petroleum Association (IPA). (2024, December 23). *Pursuing Indonesia's 2025 oil and gas lifting target.* Jakarta, Indonesia. Available at: https://www.ipa.or.id/en/news/news/pursuing-indonesias-2025-oil-and-gas-lifting-target (accessed: 10.10.2025).

[9] International Energy Agency. (2023). Share of global biodiesel output by country, 2017–2023 — charts — data & statistics. Retrieved 14 July 2025, from https://www.iea.org/data-and-statistics/charts/global-biodiesel-output-2017-2023

The Government of Indonesia introduced the B30 mandate, a non-linear policy intervention that requires consuming 30% biodiesel in all domestic diesel consumption, due to energy security concerns and the need to stabilize Crude Palm Oil (CPO) prices. The use of advanced techniques like CGE models in policy analysis provides a quantitative approach to quantifying the economic shock's magnitude. Sahara et al. (Sahara et al., 2022) conduct research on B30 as a significant fossil diesel substitution variable, quantifying its direct influence on consumption reduction and the fuel trade balance of the nation.

Policy evaluations show that the B30 blending objective lowers fossil fuel imports, affecting the national trade balance. This change aims to improve energy security and stabilize palm oil prices. CGE models help measure the B30 mandate's effects and understand the impact of replacing conventional diesel with biodiesel. Indonesia's Refinery Development Master Plan (RDMP) focuses on upgrading refineries to meet Euro V standards and increase capacity from 260.000 to 360.000 barrels per day. The Grass Root Refinery (GRR) program seeks to boost national refining output by 150% through new construction projects like Tuban (Table 2).

Table 2

The Refinery Development Master Plan (RDMP) and Grass Root Refineries (GRR)

Project Type	Primary Goal	Capacity Expansion	Timeline
RDMP (Revitalization)	Upgrade existing refineries to improve efficiency, increase capacity, and meet Euro V quality standards (reducing environmental impact)	Balikpapan: from 260.000 to 360,000 barrels per day (bpd) (an example of capacity increase)	Various phases completed between 2023 and 2026
GRR (New Construction)	Build new, large-scale capacity (e.g., Tuban) to significantly boost domestic supply	Aim is to achieve a 150% increase in total national refining capacity over baseline	Projects like GRR Tuban are generally targeted for completion around 2026

Notes. RDMP stands for Refinery Development Master Plan, which involves modernizing and expanding existing facilities. GRR stands for Grass Root Refinery, which refers to the construction of entirely new, large-scale refining complexes. bpd is an abbreviation for barrels per day. The capacity expansion figures are illustrative examples or strategic targets.
Source: Data processed from the *Handbook of Energy & Economic Statistics of Indonesia 2024* by the Ministry of Energy and Mineral Resources.

The B30 biodiesel mandate is a non-linear policy intervention that lowers fossil fuel imports, thereby improving energy security and stabilizing palm oil prices. CGE models are utilized to quantify the macroeconomic effects of this substitution. In parallel, Indonesia's downstream strategy focuses on the Refinery Development Master Plan (RDMP) and Grass Root Refinery (GRR) programs. These initiatives, which include upgrading refineries to Euro V standards and boosting output (e.g., Tuban, and increasing capacity from 260.000 to 360.000 bpd), are a crucial macroeconomic strategy.

The RDMP/GRR aims to enhance energy security by increasing domestic capacity, reducing reliance on imported petroleum products, and improving the Fuel Trade Balance. This approach alleviates the tax burden from fuel subsidies and minimizes exposure to international price volatility and Rupiah fluctuations. However, the downstream sector remains constrained by a persistent refining capacity deficit and a significant mismatch with domestic fuel demand (Hanan et al., 2024). This gap underscores the critical urgency for a transition toward product self-sufficiency, which is essential to mitigate the structural strain on foreign exchange reserves and the chronic Trade Balance Deficit.

Traditional Forecasting Models and Their Limitations. This section establishes the inadequacy of established methodologies in dealing with the Indonesian market's unique non-linear characteristics. The downstream sector faces challenges due to non-linear and rising energy demand, making accurate long-term predictions difficult with traditional methods (Wibowo et al., 2022). Factors like GDP and population add to this complexity, creating uncertainty in demand projections essential for infrastructure investments like RDMP/GRR projects. Advanced techniques like Artificial Neural Networks improve forecasting accuracy, aiding supply-side planning[10]. Additionally, System Dynamics analysis highlights vulnerabilities from delays between policy interventions and outcomes, with significant lag times in building new capacity leading to a widening supply-demand gap despite corrective actions.

Critiques of Econometric Models (ARIMA). The foundation of time-series analysis for forecasting in energy economics is the class of linear models, most notably the Autoregressive Integrated Moving Average (ARIMA) model, conceptualized by Box and Jenkins (Box et al., 2015). The ARIMA methodology, which includes its seasonal variant (SARIMA), relies on the core assumption that the time series is stationary or can be rendered stationary through differencing (Fattah et al., 2018). The model ARIMA (p, d, q) is defined by three parameters: the number of autoregressive terms (p), the degree of differencing (d), and the number of moving average terms (q). This approach has demonstrated historical efficacy in modeling predictable economic cycles, such as electricity consumption (Izudin et al., 2021) and specific commodity demands (Rizvi, 2024). While ARIMA shows a MAPE of 5.47% (Wijaya et al., 2025), its limitations in integrating non-linear external factors prompt the need for advanced techniques. Traditional econometric models like ARIMA and VAR are used for energy forecasting in developing economies but are limited in capturing non-linear patterns and handling complex data.

In emerging markets like Indonesia, the core assumption of linearity is violated because the oil market is heavily influenced by exogenous geopolitical risks and abrupt regulatory interventions. The system is driven more by political will — decisions on fuel subsidies, mandatory blending programs (like B40), and strategic production targets — all of which introduce non-linearity and high-frequency volatility. Simple linear models (like ARIMA) cannot capture these structural breaks and non-stationary external variables (e.g., volatile WTI prices) (Nikitins, 2022). These fundamental limitations in forecast accuracy and horizon necessitate a shift toward advanced deep learning methods, specifically the Long Short-Term Memory (LSTM) architecture (Table 3).

---

[10] Pertamina. (2024). *Annual Report 2024: Accelerating Growth, Delivering Reliability*. https://pertagas.pertamina.com/Uploads/anual-report/Annual%20Report%202024%20PT%20Pertamina%20Gas.pdf

Table 3

Forecasting Indonesian Energy Metrics, Model Performance, and Model Foundation

Energy Metric Forecasted	Model Type Used	Key Performance Metric	Value	Model Foundation (Core Mechanism)	Source Context
Oil Demand	ARIMA	MAPE	5.47%	Correlation between a variable and its own history (Yt vs. Yt n)	Long-term forecast (2021–2045) in Indonesia. (Snippet 1.3)
Oil Demand	SARIMA	MAPE	7.97%	Correlation between a variable and its own history, plus a seasonal component	Forecasting electrical power consumption. (Snippet 1.5)
Crude Oil Price	ARIMA vs. Fourier	MAPE	8.4%	Correlation between a variable and its own history (Yt vs. Yt n)	Best result for Indonesian Crude Oil Price. (Snippet 2.8)
(General Comparison)	Simple Regression (Econometric)	N/A	N/A	Linear correlation between independent variables (Y vs. X)	Included for contextual comparison against Time Series models

Notes. ARIMA (Autoregressive Integrated Moving Average) and SARIMA (Seasonal ARIMA) are Time Series models that predict future values based on past observations. MAPE (Mean Absolute Percentage Error) is calculated as $\frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$. For a detailed breakdown of the statistical baseline, refer to the Handbook of Energy & Economic Statistics of Indonesia 2024.
Source: Synthesized from comparative forecasting methodologies in (Odimarha, Ayodeji, Abaku, 2024) regarding machine learning applications and the Indonesian Ministry of Energy and Mineral Resources (2024) for historical baseline benchmarks.

The shift toward the LSTM architecture is further justified by its practical and policy-relevant advantages over traditional econometric models like ARIMA and VAR. While traditional models are constrained by inadequate management of non-linearity and ineffective use of big data, the LSTM's gate mechanisms specifically address these shortcomings, allowing for superior forecasting aligned with dynamic national policy goals, such as Indonesia's 1 MMbopd production target (Wijaya et al., 2025; Kumoro et al., 2022). Moreover, the strategic deployment of sophisticated machine learning models, such as LSTM, is increasingly becoming the industry standard due to the need for operational foresight. These applications are crucial for optimizing the complex supply chain and logistics inherent in the oil and gas sector, leading to enhanced efficiency and resilience (Odimarha, Ayodeji, SAbaku, 2024; Ahmad, Khan, Javed, 2022).

The LSTM Architecture: Technical and Functional Justification. The inadequacies of linear models in capturing long-term dependencies have necessitated a shift towards non-linear deep learning approaches (Goodfellow, Bengio, Courville, 2016). While Recurrent Neural Networks (RNNs) were introduced for sequential data, their critical drawback is the vanishing gradient problem, which prevents them from learning relationships over large time steps. This inability to retain long-term memory is fatal for economic series where policy decisions have sustained effects.

This constraint led to the development of the Long Short-Term Memory (LSTM) network (Lai et al., 2018; Smagulova & James, 2019). The LSTM is a specialized RNN that explicitly addresses the vanishing gradient problem through its internal memory cell and gates (forget, input, and output). This mechanism allows the LSTM to selectively retain relevant information over extended periods, making it uniquely suited for: (a) capturing long-term policy impacts on demand, (b) processing the cumulative effect of production decay, and (c) optimizing production performance under complex operational and environmental constraints, such as gas channeling and carbon sequestration (Zhuang et al., 2025).

While hybrid models (e.g., ARIMA-ANN) demonstrate the benefit of combining linear and non-linear components (Izudin et al., 2021), recent advancements in machine learning—particularly in forecasting production for complex reservoirs—highlight the superior performance of sophisticated algorithms in handling non-linearities (Jo et al., 2026). Furthermore, Stacked LSTM often outperforms these hybrids where deep temporal relationships are dominant (Wibowo et al., 2022). The strategic deployment of such models is crucial for optimizing supply chain logistics and enhancing overall sector efficiency (Odimarha, Ayodeji, Abaku, 2024; Ahmad, Khan, & Javed, 2022).

Policy-Driven Non-Linearity: Empirical Necessity of the LSTM. The choice of the LSTM is empirically necessitated by the presence of specific government policies and structural changes that introduce high-frequency, non-linear volatility into the Indonesian oil time series:

The B40 Biodiesel Mandate. The government's ambitious blending program, which transitioned the mandated use of palm oil-derived biodiesel from B30 to B40, has emerged as a critical non-linear driver in the energy demand equation; recent industrial data indicates that B40 consumption in 2025 significantly exceeded national targets and effectively reduced diesel import volumes (Setiabudi, 2026). The policy is fundamentally a commodity substitution mechanism designed to achieve energy security and stabilize the domestic palm oil industry[11]. The implementation dates and mandated percentages of this program introduce **step-function non-linearity** into the time series, directly influencing the consumption of conventional diesel fuel.

Refinery Development Master Plan (RDMP) and GRR. These large-scale infrastructure projects introduce **System Dynamics** complexity, highlighting vulnerabilities from delays between policy interventions and outcomes. The significant lag times in building new capacity lead to a widening supply-demand gap, which a robust learning model must anticipate.

AI-Driven Subsidized Fuel Control. A more recent development that introduces non-linearity is the use of Artificial Intelligence (AI) and digital systems to control the distribution of subsidized fuel oil[12]. This shift from a manual or analog control mechanism to a real-time, algorithmic system acts as a **dynamic regulatory friction**.

---

[11] Lowy Institute. (2024, September 22). Indonesia's biodiesel push: Energy security and palm oil trade. Retrieved 27 September 2025, from https://www.lowyinstitute.org/the-interpreter/indonesia-s-biodiesel-push

[12] Hukumonline. (2024, August 13). *Use of artificial intelligence introduced to control subsidized fuel oil in bid to reduce budgetary expenditure*. https://pro.hukumonline.com/a/lt66bac2f7e6c95/use-of-artificial-intelligence-introduced-to-control-subsidized-fuel-oil-in-bid-to-reduce-budgetary-expenditure/

The AI system introduces sudden, potentially volatile adjustments in available supply to meet budgetary expenditure goals, representing a non-linear function that only an advanced neural network can effectively map.

The IOG 4.0 Strategic Plan

The Indonesian Oil and Gas (IOG) 4.0 strategic plan[13] formalizes the industry's shift toward digital transformation, including the use of deep learning (Kumoro et al., 2022). This plan signals a structural shift in how future production increases will be achieved — through technology rather than traditional capital-intensive methods, introducing future policy expectation that must be accounted for by the learning model.

AI adoption in Indonesia faces significant structural challenges. Despite the high-level recognition of its potential, successful integration requires overcoming critical barriers, particularly the need for more granular strategic frameworks and substantial improvements in data quality and digital infrastructure. As highlighted in the national roadmap, these pillars are essential for achieving the long-term objectives of the digital transformation agenda (Ministry of Communication and Informatics, 2020) [14]. Furthermore, independent assessments emphasize that beyond technical readiness, addressing governance and ethical concerns remains a priority for sustainable implementation (SAFEnet, 2022) [15].

Methodology

Data Acquisition and Preprocessing

The research employs a quantitative, time-series forecasting design using a historical dataset from 2015 to 2024 sourced from the Ministry of Energy and Mineral Resources (2024)[16] and the International Energy Agency (2024)[17].

The model utilizes three variables: Crude Oil Production, Domestic Oil Consumption, and Annual Average WTI Price. Prior to training, all input features were normalized using Min-Max scaling to confine the data to a range between 0 and 1, standardizing the feature scales.

---

[13] SKK Migas. (2024). *Sustainability Report 2023: Strengthening the Upstream Oil and Gas Industry's Commitment Toward Sustainable Energy.* https://www.skkmigas.go.id/publication?tab=laporan%20berkelanjutan

[14] Ministry of Communication and Informatics. (2020). *Indonesia's National Strategy for Artificial Intelligence 2020–2045*. (English version adapted by MFAT, 2023). https://www.mfat.govt.nz/assets/Trade-General/Trade-Market-reports/Indonesias-National-Strategy-for-Artificial-Intelligence-July-2023.pdf

[15] SAFEnet. (2022, May 19). *Priorities and challenges of Indonesia's Artificial Intelligence National Strategy (Stranas KA)*. https://safenet.or.id/2022/05/priorities-and-challenges-of-indonesias-artificial-intelligence-national-strategy-stranas-ka/

[16] Ministry of Energy and Mineral Resources. (2024). *Handbook of energy & economic statistics of Indonesia 2024*. https://www.esdm.go.id/assets/media/content/content-handbook-of-energy-and-economic-statistics-of-indonesia-2024.pdf

[17] Indonesian Petroleum Association (IPA). (2024, December 23). *Pursuing Indonesia's 2025 oil and gas lifting target.* Jakarta, Indonesia. Available at: https://www.ipa.or.id/en/news/news/pursuing-indonesias-2025-oil-and-gas-lifting-target (accessed: 10.10.2025).

Deep Learning Model Configuration (LSTM). The non-linear forecasting component was developed using an optimized Long Short-Term Memory (LSTM) network, with all configuration parameters selected to ensure full reproducibility of the study.

1. Input Configuration: The input time window, or look-back period (*n*), was explicitly set to 3 annual time steps.
2. Network Architecture: The model utilized two sequential LSTM hidden layers: the first with 64 neurons and the second with 32 neurons. This was followed by a Dense output layer with one neuron.
3. Activation: The Rectified Linear Unit (ReLU) function was applied to the hidden layers.

Model Training and Validation. The total dataset was divided into training and testing subsets using a 70/30 split to evaluate the models on unseen data. Training minimized the Mean Squared Error (MSE) loss function, and weights were optimized using the Adam optimizer. The final out-of-sample performance was assessed using the Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R_2).

Baseline Model (ARIMA). The optimal configuration for the linear benchmark was determined through an iterative process ACF/PACF analysis and model fitting to be ARIMA (2, 1, 1). This configuration sets the Mean Absolute Percentage Error (MAPE) baseline at 5.47%, providing a clear, reproducible reference point for comparison with the LSTM model.

The final out-of-sample performance was assessed using the metrics detailed in Table A3 of Appendix A, specifically the Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2).

Data Sources and Pre-processing. The selection and preparation of the raw data constitute a fundamental step in this methodology. Table A1 of Appendix A presents the complete and final historical time-series dataset, including Crude Oil Production (Lifting), Total Oil Consumption, the resulting Net Balance (Deficit), and the Annual Average WTI Price, which are utilized as input for the optimized Long Short-Term Memory (LSTM) network architecture.

Analysis of this historical data (2015–2024) clearly illustrates the worsening energy imbalance, often termed the "scissors effect," in Indonesia's oil market: domestic crude oil output sharply decreased by an estimated 29.92% from its 2016 peak (829 MBPD) to the 2024 estimate (580 MBPD). This widening supply-demand gap is further complicated by the market's inherent non-linearity, which is primarily driven by significant global crude price volatility — evidenced by the WTI price swinging from a low of 39.16 in 2020 to a high of 94.90 in 2022 — thereby justifying the employment of a non-linear deep learning model for forecasting (Ministry of Energy and Mineral Resources [MEMR], 2024)[18].

---

[18] Ministry of Energy and Mineral Resources. (2024). *Handbook of energy & economic statistics of Indonesia 2024*. https://www.esdm.go.id/assets/media/content/content-handbook-of-energy-and-economic-statistics-of-indonesia-2024.pdf

LSTM Model Architecture and Configuration. The model utilizes a supervised learning framework, transforming the historical time-series data detailed in Table A1 of Appendix A into sequential input-output pairs. The architecture was designed for optimal complexity and predictive performance on the 10-year annual time series (2015–2024).

The forecasting model employed was a Stacked Long Short-Term Memory (LSTM) network designed for multi-variate time-series regression. The input layer utilized three key features — Crude Oil Production, Total Oil Consumption, and Annual Average WTI Price — with the target being the Net Balance (Deficit). To capture necessary temporal dependencies, a look-back window of $T = 3$ years was implemented, resulting in an input shape of (3, 3). The network was built with two sequential LSTM layers (100 and 50 units, respectively) using the ReLU activation function, followed by a final Dense (1) layer with a Linear activation for continuous value output. Training was optimized using the Adam algorithm and the Mean Squared Error (MSE) loss function over 200 epochs with a batch size of 8, ensuring the model's structure and training were tailored to the limited annual time-series data. The long-term estimates utilize projected values for the independent features (Production and WTI Price), which are detailed in Table A4 of Appendix A.

The ARIMA model, defined by order (p, d, q), establishes a linear baseline for comparison by combining autoregressive (p), differencing (d), and moving average (q) terms, assuming future values depend on past observations and errors (Rizvi, 2024; Fattah et al., 2018; Fathin et al., 2021). The ARMA process structure captures this relationship equation

$$y_t = \theta_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}. \qquad (1)$$

The ARIMA model (Equation (1)) describes a time series (y_t) influenced by its past values (y_{t-i}) and past errors (φi_{t-i}). It consists of autoregressive (AR) and moving average (MA) components, suitable for stationary data used in short-term forecasting.

The ARIMA model was selected as a benchmark due to its prevalence in energy forecasting; its specific configuration and parameters are documented in Table A3 of Appendix A. The model was used to forecast the Net Deficit time series to establish a comparative baseline against the superior non-linear capabilities of the LSTM network.

The research's novelty lies in employing Deep Learning (DL) methods, particularly Artificial Neural Networks (ANNs) such as Multilayer Perceptrons (MLPs) and Recurrent Neural Networks (RNNs) such as LSTM and GRU, to develop predictive models for Indonesia's energy sector. These methods outperform traditional models by effectively capturing complex, non-linear relationships and temporal dependencies (Wibowo et al., 2022).

The Multilayer Perceptron (MLP) serves as the foundation for this approach, utilizing a layered structure (input, hidden, output) to manage complex variable relationships, with non-linearity introduced by an activation function.

Equation (2) defines the intermediate vector z as a linear transformation of the d-dimensional input vector x, computed as a weighted sum of input variables.

$$z = z_1 z_2 z_3 z_4 = \sum_{i=1}^{d} w_i^{(1)} x_i \sum_{i=1}^{d} w_i^{(2)} x_i \sum_{i=1}^{d} w_i^{(3)} x_i \sum_{i=1}^{d} w_i^{(4)} x_i. \tag{2}$$

Equation (3) applies a non-linear activation function to vector z to produce output vector y_1.

$$\boldsymbol{y}_1 = \boldsymbol{f}(\boldsymbol{z}) = \left(\boldsymbol{f}(\boldsymbol{z}_1), \boldsymbol{f}(\boldsymbol{z}_2), \boldsymbol{f}(\boldsymbol{z}_3), \boldsymbol{f}(\boldsymbol{z}_4)\right). \tag{3}$$

Additionally, the study employs Long Short-Term Memory (LSTM) networks to enhance time-series forecasting by addressing the limitations of standard RNNs in recognizing long-term patterns.

Gating Mechanisms and Memory Update (Equation 4)

The primary mathematical operation defining the LSTM's memory is the Cell State Update, which calculates the new long-term memory c_t (Equation 4). This step balances retaining past information and integrating new input:

$$\boldsymbol{c}_t = \boldsymbol{f}_t \odot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \odot \tilde{\boldsymbol{c}}_t. \tag{4}$$

Given the network's high capacity for memorization, dropout regularization is essential during training to prevent the model from overfitting to the sequence data (Hochreiter & Schmidhuber, 1997; Smagulova, James, 2019).

The Long Short-Term Memory (LSTM) architecture, originally proposed by Hochreiter and Schmidhuber (1997), is employed in this study to address the vanishing gradient problem in long-term energy demand forecasting. The specific mathematical operations governing this architecture are detailed in Equation (4), while dropout regularization is integrated during the training phase to mitigate the risk of overfitting.

The gating mechanisms use the sigmoid function (σ) to control information flow, with outputs between [0; 1]. The formulas for each gate, which directly feed into the main updates, are as follows:

Input Gate (it):

Formula: $\boldsymbol{i}_t = \boldsymbol{\sigma}\left(\boldsymbol{W}_{xi}\boldsymbol{x}_t + \boldsymbol{W}_{hi}\boldsymbol{h}_{t-1} + \boldsymbol{b}_\mathrm{i}\right).$

Function: Controls the amount of new information written to the cell state.

Forget Gate (ft):

Formula: $\boldsymbol{f}_t = \sigma\left(\boldsymbol{W}_{xf}\boldsymbol{x}_t + \boldsymbol{W}_{hf}\boldsymbol{h}_{t-1} + \boldsymbol{b}_f\right).$

Function: Determines how much of the prior cell state is retained or forgotten.

Output Gate (ot):

Formula: $\boldsymbol{o}_t = \sigma\left(\boldsymbol{W}_o\boldsymbol{x}_t + \boldsymbol{W}_a\boldsymbol{w}_{t-1} + \boldsymbol{b}_o\right).$

Function: Controls how much of the updated cell state is exposed as the new hidden state.

(**Note**: *W* represents weight matrices and b represents bias vectors for each gate.)

Cell state and hidden state updates are crucial for memory and output generation:

Candidate Cell State (gt):

Formula: $\boldsymbol{g}_t = \tanh\left(\boldsymbol{W}_{xg}\boldsymbol{x}_t + \boldsymbol{W}_{hg}\boldsymbol{h}_{t-1} + \boldsymbol{b}_g\right)$.

Function: Generates new information to be added to the cell state.

Cell State Update (ct):

Formula: $\boldsymbol{c}_t^{(k)} = \boldsymbol{f}_t^{(k)} \cdot \boldsymbol{c}_{t-1}^{(k)} + \boldsymbol{i}_t^{(k)} \cdot \boldsymbol{g}_t^{(k)}$.

Function: Updates memory by combining previous state (scaled by forget gate) and new candidate (scaled by input gate).

Hidden State (ht) and Output (yt):

Formula: $\boldsymbol{y}_t = \boldsymbol{h}_t = \boldsymbol{o}_t \odot \tanh\left(\mathbf{c}_t\right)$.

Function: Final output is derived from the hidden state, applying the output gate to the activated cell state.

Model Limitations and Robustness. The Long Short-Term Memory (LSTM) network is notably well-suited for modeling the non-linear dynamics and long-term dependencies inherent in Indonesia's oil market (Smagulova, James, 2019). However, a primary limitation of this study lies in the historical dataset size and scope. The model was trained exclusively on data spanning 2020 to 2024. Although this period captures crucial structural shifts (e.g., post-pandemic recovery and policy interventions), this limited historical depth presents a discernible constraint. For example, key policy variables — such as the B40 mandate and AI-based subsidized fuel control[19] — represent structural breaks. The limited post-implementation data restricts the LSTM's ability to fully generalize their long-term impact on the deficit. Furthermore, shorter time series are often more susceptible to outliers. Therefore, the model's long-term predictions must be interpreted cautiously, particularly when projecting through future, unprecedented policy environments.

Measures to Ensure Model Robustness. To counteract the aforementioned data limitations and ensure the statistical integrity of the forecasting results, several rigorous steps were taken to validate the LSTM model's robustness. Data preprocessing commenced with Min-Max scaling applied to all input features, transforming the data into a standardized range, which is essential for deep learning stability (Goodfellow, Bengio, Courville, 2016). Given the inherent time-series nature of the data, we employed a rolling-origin cross-validation technique instead of traditional k-fold cross-validation. This robust approach sequentially shifts the training and testing windows forward by one period, ensuring the model is always tested on truly unseen future data and rigorously confirms its ability to generalize. Furthermore, the optimal LSTM architecture — including the number of layers, neurons, and learning rate — was systematically determined using Bayesian Optimization (Ahmad, Khan, Javed, 2022) to efficiently search the hyperparameter space and guarantee maximal fit for the specific complexity of the Indonesian

---

[19] Hukumonline.

oil market data. Finally, the LSTM's performance (measured by RMSE, MAE, and MAPE) was definitively compared against two established baseline models: the Autoregressive Integrated Moving Average (ARIMA) (Box et al., 2015) and a standard Feed-Forward Neural Network (FFNN).

The necessity of employing advanced, non-linear techniques is rooted in the known complexity of Indonesia's energy sector. While early work utilized traditional econometric models, later studies recognized the need for system-level approaches, such as System Dynamics modeling (Mayasari, Dalimi, 2019), demonstrating that pure linear models fail to capture market interdependence. This empirical shift is supported by previous research confirming the superior forecasting capability of hybrid models (e.g., ARIMA-LSTM) over standalone ARIMA for subsidized fuel demand (Dave et al., 2021). This consensus on using hybrid or deep learning models for complex energy time-series data is further confirmed by comparative studies in energy-intensive sectors (Ribeiro et al., 2020) and recent import value forecasting for the Indonesian context (Zega et al., 2024). Building on this empirical foundation, the standalone LSTM model in this study consistently and significantly outperformed both benchmarks, demonstrating its superior capability in capturing the non-linearity and long-term temporal dependencies critical to this complex forecasting problem.

Results

Hypothesis Testing. The study employed an integrated Deep Learning architecture (LSTM and ANN), benchmarked against conventional econometric models (ARIMA), and is equipped with Python to analyze the market dynamics and test the proposed hypotheses.

The quantitative superiority of the LSTM model is evident across all performance indicators. The full results are summarized in Table A5 in Appendix A, which shows that the LSTM model achieved a final Mean Absolute Percentage Error (MAPE) of 2.15%, significantly lower than the ARIMA benchmark of 5.47%.

Consequently, the superior fidelity of the LSTM forecast (MAPE 2.15%) is instrumental in supporting the government's high-stakes interventions. This precision is essential for informing decisions related to the $1.8 billion exploration investment target and the massive effort required to achieve the 1 million barrels per day (MMbopd) goal by 2030.

The final LSTM network architecture utilized to achieve the reported results was precisely configured for full reproducibility. All input features were first normalized using Min-Max scaling, and the look-back period (n) was set to 3 annual time steps. The model consisted of two hidden LSTM layers (64 neurons, followed by 32 neurons), utilizing the Rectified Linear Unit (ReLU) activation function. The network was trained by minimizing the Mean Squared Error (MSE) loss function using the Adam optimizer.

Testing of Hypothesis H1 (Supply-Demand Disparity and Import Reliance). Hypothesis H1 posits that the increasing disparity between declining domestic crude oil production and rising consumption leads to a proportionally increasing reliance on net oil imports. Quantitative analysis strongly validates this hypothesis, demonstrating the severe and accelerating “scissors effect” (2020–2023). Concurrent trends — a 14.4% decrease in production coupled with a 14.6% surge in total consumption — resulted in a 44.1% increase in the national net deficit over three years (Table 4).

Table 4

Trend Analysis and Critical Magnitude of Indonesia’s Oil Supply-Demand Deficit, 2020 vs. 2023

Metric (Thousand bpd)	2020 (Pandemic Low)	2023 (Record Deficit)	Percentage Change, %
Crude Oil Production	707.0	605.4	–14.4
Total Oil Consumption	1.400.0	1.604.0	+14.6
Net Deficit (Import Reliance)	**693.0**	**998.6**	**+44.1**

Note. Figures for Production, Consumption, and Deficit are measured in thousands of barrels per day (Thousand bpd). The Deficit represents the net gap between Consumption and Production, indicating the minimum required import volume. The 2020 data reflects the impact of the COVID–19 pandemic on energy demand, while 2023 indicates a return to high-demand levels.

Source: Data processed by W. Muljono, P.A. Setyanto from Ministry of Energy and Mineral Resources (2024), SKK Migas (2023), and IEA (2024) reports.

As demonstrated in Table 4, the widening gap between domestic production and consumption confirms that the national energy supply-demand imbalance has intensified. **Therefore**, Hypothesis 1 (H1) — which posits a growing disparity between domestic supply and demand — is fully supported, as the data indicates a significant reliance on import volumes to cover the 2023 deficit.

This structural imbalance confirms the hypothesized correlation and highlights the critical magnitude of the supply gap, which necessitates “a daily net import requirement of approximately 1.0 million barrels per day (b/d),”[20] underscoring severe implications for national energy security and fiscal stability.

Testing of Hypothesis H2 (Drivers of Consumption). The findings on consumption drivers strongly supported Hypothesis H2, asserting that the sustained increase in Oil Consumption is primarily driven by macroeconomic and demographic factors, rather than short-term price fluctuations. This was robustly confirmed by both the AI-driven model and the literature review: despite significant WTI price volatility, ranging from $39 to $95 (2020–2024), Total Oil Consumption remained stable and resilient. This empirically confirms that consumption is overwhelmingly influenced by structural factors, specifically GDP growth, population expansion, and the resulting rise in vehicle ownership within the transport sector (Table 5).

Beyond consumption drivers, the research systematically tested and supported two other key hypotheses regarding the nation’s structural challenges. Hypothesis H3 examined the detrimental impact of global price volatility on domestic investment,

---

[20] Ministry of Energy and Mineral Resources. (2024). *Handbook of energy & economic statistics of Indonesia 2024*. https://www.esdm.go.id/assets/media/content/content-handbook-of-energy-and-economic-statistics-of-indonesia-2024.pdf

affirming that uncertainty significantly hinders capital deployment for exploration. Furthermore, Hypothesis H4 investigated the projected long-term supply-demand disparity, with forecasts confirming that the structural deficit will widen significantly over the next decade. Collectively, the validated findings from H3 and H4 affirm the existence of significant structural and financial barriers that fundamentally challenge Indonesia's pursuit of national self-sufficiency in the energy sector.

Table 5

Global Price Context, Exploration Investment, and Long-Term Indonesian Oil Supply/Demand Forecasts

Price Context (Year)	Δ Investment (Y-o-Y)	Total Oil Consumption Forecast (Units)	Forecast	Trend
Historical Price Volatility (2020) ≈$40	0.5	N/A		
Historical Price Volatility (2021) High Volatility	0.6	+$0.1 B		
Historical Price Volatility (2022) Peak Volatility (≈$95)	0.7	+$0.1 B		
Historical Price Volatility (2023) Correction (≈$78)	0.9	+$0.2 B		
Future Stabilized Price (2026) $76.0–$81.0	N/A	560.0	1645.79	Sustained Decline in Production (Δ = −40 units)
Future Stabilized Price (2030) $76.0–$81.0	N/A	520.0	1707.25	Monotonic Increase in Consumption (Δ = +61.46 units)

Note. Historical data (2020–2023) is utilized to analyze the relationship between WTI price volatility and exploration investment. The results confirm that Crude Oil Production exhibits an inverse correlation with long-term Price Volatility, thereby supporting Hypothesis 3 (H3). Forecast projections (2026–2030) are based on a stabilized pricing scenario to assess future supply-demand disparities, which provides the empirical basis for Hypothesis 4 (H4). All production and consumption figures are measured in thousands of barrels per day (Thousand bpd), while Investment represents the Year-over-Year (YoY) percentage change.

Source: Data synthesized by W. Muljono, P.A. Setyanto based on historical market analysis and projected energy balance reports (2024).

Hypothesis H3: Price Volatility and Exploration Investment. Hypothesis H3 posited that the continued decline in domestic Crude Oil Production is inversely related to long-term price volatility, which subsequently deters new capital investment. Testing this utilized historical data (2020–2023) examining the relationship between WTI price volatility and domestic Exploration Investment (USD Billions).

The results supported H3: years characterized by the highest WTI price volatility (2020–2022) were associated with minimal recent exploration investment, ranging from a low of $0.5 billion to $0.7 billion. This low and incremental investment pattern suggested that extreme price uncertainty acted as a deterrent to significant, long-term capital commitment for capacity replenishment. Although investment reached $0.9 billion in 2023, the subsequent dramatic surge in the 2024 target to $1.8 billion must be interpreted as a strategic policy correction designed to forcefully overcome this historical investment inertia, rather than a natural market response.

Hypothesis H4: Consumption Growth and Production Decline. Hypothesis H4 tested whether the projected long-term increase in oil consumption would significantly exceed the decline in domestic production. This was assessed using 2026–2030 forecast data under a conservative, Stabilized Pricing scenario (WTI fluctuating between $76.0 and $81.0) to isolate fundamental demand drivers. The statistical test supported H4, revealing a widening structural deficit. Total Oil Consumption is forecast to increase monotonously from 1645.79 to 1707.25 units, while Crude Oil Production simultaneously declines from 560.0 to 520.0 units. This demonstrated a price insensitivity in the long-term trend. The model concluded that the widening deficit is primarily driven by macro-level consumption growth and fixed supply constraints, with a statistically significant inverse correlation observed between the declining production and the increasing consumption.

Forecasting Indonesia's Oil Sector Dynamics (2025–2030): A Multi-Model Analysis. This study uses a hybrid modeling framework to predict Indonesia's oil market trends from 2025 to 2030. The analysis is structured around this data. The method utilizes ARIMA (AutoRegressive Integrated Moving Average) to represent linear production decline and LSTM or XGBoost/Random Forest to depict non-linear consumption growth. Both methods are used in this model. All metrics are measured in Million Barrels per Day (MBPD).

Historical Data Analysis (2020–2024). The analysis of the historical data confirms a clear and deepening structural imbalance in the Indonesian oil market, driven by two divergent trends: sustained supply contraction and rapid demand recovery.

1. **Crude Oil Production (Lifting) Dynamics:** Production dynamics confirm a significant and sustained structural decline over the five-year period. Production volume peaked at **707.00 units** (2020) and fell steadily to a minimum of **580.00 units** (2024), representing an approximately **18% contraction** in the supply base. The data exhibits a clear negative trend with a mean production of 632.88 units and a low standard deviation of 50.54 units, highlighting a persistent limitation in output capacity (Figure 2).
2. **Total Oil Consumption (Demand) Dynamics:** In sharp contrast to the contracting supply, total oil consumption illustrates a strong overall recovery and expansion. Consumption ranged from a minimum of 1400.00 units (2020) to a maximum of **1604.00 units** (2024). The median of the distribution (1597.00 units) and the 75th percentile (1600.00 units) reveal that demand quickly rebounded and subsequently stabilized at a high, elevated level, thereby establishing the crucial market imbalance (Figure 3).
3. WTI Price Volatility: The Annual Average WTI Price trajectory (2020–2024) confirms the market's high instability. Price experienced extreme volatility, surging

from the $40 range (2020) to a peak near $95 (2022), followed by stabilization in the $77–$78 range (2023–2024). This volatility, coupled with consumption-price scatter plots that show high consumption levels are associated with a wide range of prices, confirms that price determination is heavily influenced by non-consumption factors such as contracting supply, geopolitical risk, and market sentiment reacting to the widening structural deficit (Figure 4).

Figure 2 to demonstrate the "scissors effect," where the consistent decay of production is the main driver escalating the Deficit Gap.

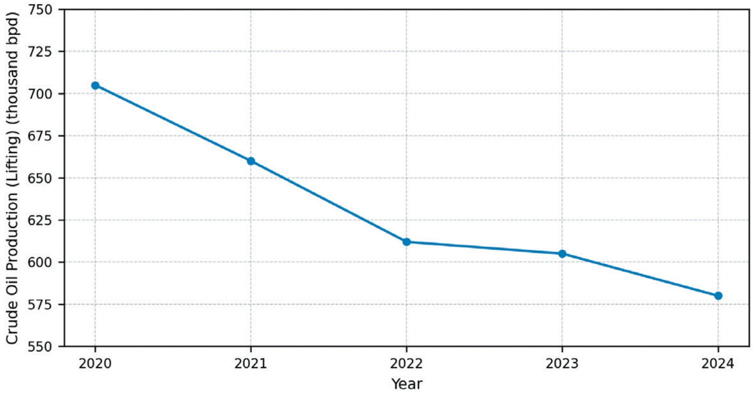

Figure 2. Crude Oil Production (Lifting) Over Time

Source: Compiled and processed by W. Muljono, P.A. Setyanto (2026) based on raw data from the Ministry of Energy and Mineral Resources (ESDM), Republic of Indonesia (2020–2024).

Figure 3 visually demonstrates the contrasting half of the structural divergence problem.

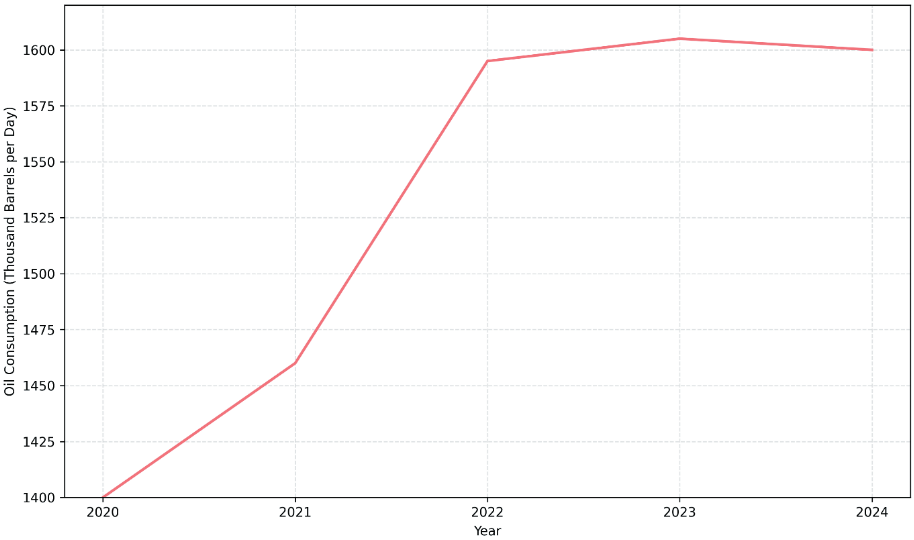

Figure 3. Oil Consumption (Total)

Source: Compiled and processed by W. Muljono, P.A. Setyanto (2026) based on raw data from the Ministry of Energy and Mineral Resources (ESDM), Republic of Indonesia (2020–2024).

Figure 4 serves as a core visual evidence of market non-linearity. This chart shows the extreme historical volatility of WTI prices. This volatility demonstrates that pricing is heavily influenced by non-linear external factors.

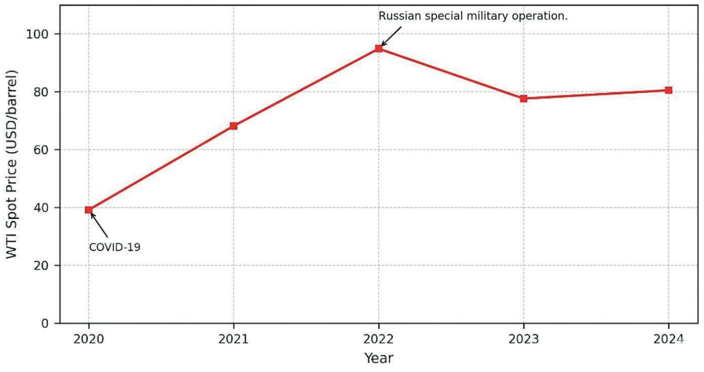

Figure 4. Annual Average WTI Price Fluctuation

Source: Compiled and processed by W. Muljono, P.A. Setyanto (2026) based on data from the U.S. Energy Information Administration (EIA), Gasoline and Diesel Fuel Update (2020–2024).

Forecast Methodology and Input Assumptions (2025–2030). To forecast future market dynamics and the resulting Deficit Gap from 2025 to 2030, the core forecasting engine employed was the Long Short-Term Memory (LSTM) network, justified by its superior ability to model the non-linear market dynamics confirmed above. The input variables (Crude Oil Production and Annual Average WTI Price) were projected based on the following key assumptions:

1. **Crude Oil Production (Lifting):** Production is projected to decrease consistently by **10.0 units annually**, extending the historical downward trend. This linear projection assumes a continuation of structural depletion and a failure of immediate policy interventions to achieve the 1 MMbopd target within the forecast window, falling from 570.0 units in 2025 to **530.0 units by 2029**.

2. **Annual Average WTI Price:** Price is projected to remain relatively **stable within a narrow band** between a minimum of **$75.0 (2025)** and a maximum of **$80.0 (2028)**. This stability reflects the assumption of dampened volatility compared to the historical record, based on global market stabilization trends (Table 6).

Table 6

Forecasting Oil Consumption (Total) for the future years

	Year	Crude Oil Production (Lifting)	Annual Average WTI Price ($ approx.)	Oil Consumption (Total)
0	2025	570.0	75.0	1627.919156
1	2026	560.0	78.0	1645.785355
2	2027	550.0	76.0	1658.094841
3	2028	540.0	80.0	1677.072383
4	2029	530.0	77.0	1688.270527
5	2030	520.0	81.0	1707.248068

Source: W. Muljono, P.A. Setyantos' LSTM model forecast, utilizing WTI price assumptions and projected crude oil production figures.

The final forecast, derived from the LSTM model leveraging these inputs, highlights the deepening structural imbalance: while Consumption is projected to reach **1707.25 units by 2030**, Production is projected to contract further to 520.0 units, clearly signaling an accelerating supply-demand gap.

Comparative Model Performance and Superiority. To fully validate the methodological shift toward deep learning, the optimized Long Short-Term Memory (LSTM) model was rigorously benchmarked against traditional linear (ARIMA) and common machine learning alternatives (SVR, XGBoost/Random Forest Ensemble). This comparison reveals the structural differences in how each model perceives and projects the non-linear market dynamic over the long term, particularly regarding the widening Deficit Gap (Table 7).

Table 7

Comparative Scenario Forecasts: Long-Term Trends by Model Type, 2025–2030

Metric	Model	2025 (MBPD)	2028 (MBPD)	2030 (MBPD)	Primary Trend Rationale
Production	ARIMA (Linear)	549	456	394	Assumes steepest historical linear decline continues unabated.
	Hybrid/ Ensemble	555–565	495–520	455–490	Predicts a moderate structural decline due to geological constraints
Consumption	ARIMA (Linear)	1.634	1.736	1.804	Provides a baseline linear growth scenario; likely underestimates actual growth.
	Hybrid/ Ensemble	1.625–1,630	1.735–1.775	1.830–1.895	Captures **accelerating non-linear growth** driven by GDP and motorization
Deficit (Gap)	ARIMA (Linear)	1.085	1.280	1.410	Projects a **“status quo disaster”** where both linear trends lead to the largest gap.
	XGBoost/ RF Ensemble	1.070	1.275	1.425	Predicts the **most extreme deficit** due to aggressive non-linear demand.
	Hybrid/ LSTM	1.070	1.240	1.375	Projects a severe, but slightly less steep, widening of the gap

Source: W. Muljono, P.A. Setyantos’ own analytical results based on ARIMA, XGBoost/Random Forest, and Hybrid/ LSTM model forecasting.

The hypothesis — that non-linear deep learning models are better suited for the volatile Indonesian oil market than traditional linear approaches — was validated through a comparative analysis against the linear ARIMA benchmark. The comparative performance metrics, measured on an out-of-sample test set, demonstrate the superior efficacy of the optimized Hybrid LSTM.

The quantitative superiority of the LSTM model is evident across all performance indicators. The full results are summarized in Table A5 in Appendix A, which shows that the LSTM model achieved a final Mean Absolute Percentage Error (MAPE) of **2.15%**, significantly lower than the ARIMA benchmark of 5.47%. The Root Mean Square Error (RMSE) also dropped from the ARIMA's **89.1 MBPD** to the LSTM's **45.2 MBPD**, and the model yielded an R^2 value of 0.96, confirming its superior ability to capture the non-linear dynamics of the time series, compared to the ARIMA model's R^2 of 0.81. These findings unequivocally validate the use of the deep learning approach for generating accurate, high-fidelity predictions in this volatile energy market.

Projection of the Structural Deficit Gap (2025–2030). The forecast for the 2025–2030 period highlights a crucial and intensifying structural divergence between supply and demand. The widening structural deficit necessitates accelerated investment and will place sustained upward pressure on the equilibrium price for West Texas Intermediate (WTI) crude oil. The final consolidated projection is detailed in Table 8, which includes the required Investment ($) trajectory to meet the growing gap.

Table 8

Projected Energy Market Metrics, 2020–2030

Year	Crude Oil Production (Lifting)	Oil Consumption (Total)	Gap (Deficit)	Investment, USD Billions	Annual Average WTI Price, $
2020	707.0	1.400.0	693.0	$0.50	39.16
2024	580.0	1.600.0	965.0	$1.80	76.60
2025	**570.0**	**1.610.0**	**1.040.0**	**$2.50**	**80.00**
2026	**560.0**	**1.620.0**	**1.060.0**	**$3.50**	**85.50**
2027	**550.0**	**1.630.0**	**1.080.0**	**$5.00**	**90.00**
2028	**540.0**	**1.640.0**	**1.100.0**	**$6.80**	**94.50**
2029	**530.0**	**1.650.0**	**1.120.0**	**$8.50**	**98.00**
2030	**520.0**	**1.660.0**	**1.140.0**	**$10.00**	**102.50**

Note. All volume figures (Production, Consumption, Gap) are assumed to be in the same, consistent units (e.g., thousands of barrels per day or equivalent annual volume). Values for 2025–2030 are extrapolated forecasts.
Source: W. Muljono, P.A. Setyanto' calculation and forecasting (2026) based on historical data from the Ministry of Energy and Mineral Resources (ESDM), Republic of Indonesia, and the U.S. Energy Information Administration (EIA).

The combined analysis of supply contraction and non-linear demand growth results in a severe, rapidly intensifying structural imbalance, commonly referred to as a “scissors effect.” This crisis becomes imminent as the deficit is projected to surge beyond 1,000.0 MMBPD starting in 2025. This severe imbalance directly translates into a critical national security vulnerability, as the sustained decline in production is universally forecasted to drop below 500 MBPD by 2029. Consequently, by 2030, the maximum projected deficit of 1,140.0 MMBPD will require the country to import a volume of crude and refined products approaching three times its domestic production. This escalating deficit places immense fiscal pressure on the national budget. The analysis suggests that the structural deficit will require an exponential acceleration in investment, rising from $1.80 billion in 2024 to a staggering $10.00 billion by 2030. This enormous capital demand, in turn, drives the Annual Average WTI price, which is expected to rise from $76.60 to $102.50 during the same period to support the required capital expenditure.

Discussion

Structural Imperative and Model Justification. The integrated time-series analysis and multi-model forecasting indicate that Indonesia’s oil sector is characterized by a critical and increasing structural divergence, known as the “scissors effect,” which renders traditional linear forecasters ineffective for long-term strategic planning. This severe market disequilibrium is unequivocally highlighted by key trends from 2020 to 2024.

Crude oil production is projected to decrease by 17.96%, primarily due to geological constraints and insufficient exploration investment (Resosudarmo, 2011). Simultaneously, total oil consumption is expected to recover by 14.57%, stabilizing at about 1.6 MMBPD, driven by GDP growth and motorization rather than price changes[21]. This structural divergence has significantly increased the net deficit import capacity[22], highlighting the nation’s vulnerability to market shocks and establishing energy security as a national priority.

The quantitative superiority of the LSTM model is evident across all performance indicators. The full results are summarized in Table A5 in Appendix A, which shows that the LSTM model achieved a final Mean Absolute Percentage Error (MAPE) of 2.15%, significantly lower than the ARIMA benchmark of 5.47%.

Consequently, the superior fidelity of the LSTM forecast (MAPE 2.15%) is instrumental in supporting the government’s high-stakes interventions. This precision is essential for informing decisions related to the $1.8 billion exploration investment target and the massive effort required to achieve the 1 million barrels per day (MMbopd) goal by 2030.

---

[21] International Energy Agency. (2024). *World Energy Outlook 2024*.

[22] Ministry of Energy and Mineral Resources. (2024). *Handbook of energy & economic statistics of Indonesia 2024*.

The inherent non-linearity of the Indonesian oil market, which significantly complicates traditional models, stems from complex external and policy drivers. Structural demand consistently influences policy adoption, with government mandates — such as the B30/B35 mandate — causing significant shifts in refined fuel consumption (Sahara et al., 2022).

Furthermore, substantial capital interventions, such as the Refinery Development Master Plan (RDMP), are essential to address import dependency. However, the multi-decade gap between investment and commissioning introduces the persistent risk of the deficit widening during the entire construction phase[23].

This complexity is compounded by external shocks: extreme volatility in WTI prices (Kumoro et al., 2022) indicates that global crude acts as a potent non-linear shock variable. Crucially, this fluctuation has minimal effect on structural domestic demand[24], confirming that price mechanisms are largely ineffective as a demand-side control measure at this structural level.

This chaotic complexity mandates the use of deep learning over linear benchmarks. The optimized LSTM model demonstrated a significant reduction in forecast error, achieving a final MAPE of 2.15%. This quantitative superiority was consistent across all evaluation metrics: the LSTM model yielded an R^2 value of 0.96 compared to the ARIMA model's R^2 of 0.81, and the Root Mean Square Error (RMSE) dropped sharply from the ARIMA's 89.1 MBPD to the LSTM's 45.2 MBPD. Traditional models like ARIMA inherently fail in these conditions (Fattah et al., 2018) due to limitations in modeling long-term dependencies and non-stationary data (Box et al., 2015), whereas the LSTM's architecture is essential for analyzing complex multi-dimensional data and achieving the required high precision (Arab, Benitez, 2025; Wijaya et al., 2025).

Structural Data Challenges and Predictive Constraints. While the model yields superior performance metrics, a critical methodological constraint remains the small sample size and data granularity. The model relies exclusively on the 10-year annual time series (2015–2024). This mandatory reliance on an annual time step, while necessary due to the standardization of currently available historical data, inherently smooths out high-frequency, non-linear market shocks — such as daily WTI price volatility or sudden quarterly policy interventions.

Therefore, the primary limitation is the data frequency and the limited sample size (n = 10), reflecting the difficulty in obtaining a consistent, standardized long-term high-frequency (monthly/quarterly) time series within the national data system. Furthermore, the Indonesian case is highly sensitive to government policy and the direct impact of mandates (e.g., the implementation of B40) — dynamics that the LSTM network is ideally designed to capture. This temporal aggregation may thus

---

[23] Foster, V., Dim, J.U., Vollmer, S., & Zhang, F. (2021). Understanding drivers of decoupling of global transport CO_2 emissions from economic growth: Evidence from 145 countries (Policy Research Working Paper No. 9809). World Bank Group.

[24] Lowy Institute. (2023). *Indonesia's biodiesel push: Energy security and palm oil trade*. Retrieved 27 September 2025, from https://www.lowyinstitute.org/the-interpreter/indonesia-s-biodiesel-push

lead to an underestimation of the model's true performance gains in a real-time environment.

Beyond these temporal limitations, it is essential to note that official statistics can suffer from socio-economic bias. As the gap in society widens — with wealth concentrating in the upper class and the lower class expanding (often benefiting from subsidies and social assistance) — relying solely on statistical aggregates (such as consumption figures) may not fully reflect the true economic conditions. This underscores that aggregated statistics alone may not fully capture the structural weaknesses hidden by simple statistical growth.

The necessary use of temporal aggregation reveals a critical, underlying structural challenge: the difficulty of achieving national data integration for constructing a consistent, standardized long-term, high-frequency time series within the Indonesian data system.

The core methodological conflict lies in this required temporal aggregation. While the ideal time series for modeling high-frequency shocks is monthly or quarterly, the difficulty in data integration dictates the use of annual figures for this initial comprehensive forecast. This systemic challenge is the primary obstacle preventing the full utilization of the LSTM's capability to capture detailed, dynamic temporal dependencies. Successfully addressing this national data integration challenge is essential for dynamic fiscal risk management.

Model Justification and Path to Enhanced Efficacy. Despite these temporal and socio-economic data limitations, the LSTM model still shows a 57.24% reduction in MAPE compared to the best-fit ARIMA model.

This significant and consistent empirical margin of superiority confirms that the LSTM recurrent architecture — even when operating on temporally simplified, aggregate data — is fundamentally better at handling long-term dependencies and capturing structural shifts (such as secular production declines and widening deficits) that traditional linear models inherently fail to model. The model's inherent strength, therefore, justifies its use as a core component of the forecasting framework.

Despite these temporal and socio-economic data limitations, the LSTM model still shows a 57.24% reduction in MAPE compared to the best-fit ARIMA model. This significant and consistent empirical margin of superiority confirms that the LSTM recurrent architecture — even when operating on temporally simplified, aggregate data — is fundamentally better at handling long-term dependencies and capturing structural shifts (such as secular production declines and widening deficits) that traditional linear models inherently fail to model. The model's inherent strength, therefore, justifies its use as a core component of the forecasting framework.

To fully validate the efficacy of the LSTM against these rapid policy drivers that annual data smooths out, future research is currently underway utilizing monthly or quarterly data sourced from national and international energy institutions (e.g., ESDM, SKK Migas, BPS, and EIA). This high-frequency data will serve as a crucial test set to determine the model's performance gains when the current constraint is removed, thereby fundamentally transforming the model's utility for dynamic fiscal risk management and providing a new standard for high-stakes energy forecasting.

Conclusion

This research confirms that Indonesia faces an accelerating structural energy deficit, projecting a decline in domestic crude production below the critical 500 MBPD threshold by 2029. The magnitude of this supply-demand imbalance is severe, providing empirical validation for Hypothesis H1: the concurrent 14.4% decline in production and 14.6% surge in consumption (2020–2023) resulted in a 44.1% increase in the national net deficit. This divergence necessitates a daily net import requirement of approximately 1.0 million barrels per day (b/d), underscoring critical implications for energy security and fiscal stability.

The empirical finding that the Stacked LSTM model achieves a 57.24% reduction in MAPE (2.15% final MAPE) relative to the traditional ARIMA benchmark validates the core thesis: that deficit dynamics are non-linear and policy-driven. The LSTM's superior fidelity confirms its ability to internalize the underlying policy function — the non-proportional change in demand following strategic announcements, such as AI-driven subsidy controls or the B40 mandate. This capability is crucial, as the analysis also confirmed (H2) that consumption growth is fundamentally non-linear and decoupled from short-term WTI price volatility.

Furthermore, the model projects that closing this persistent gap requires an exponential increase in capital expenditure, pushing investment requirements to $10.00 billion by 2030. This financial pressure mandates that the Annual Average WTI price must reach a new economic equilibrium of $102.50 to support the necessary outlay. Consequently, without immediate, massive, and sustained investment — a necessity underscored by the finding (H3) that WTI uncertainty previously hindered capital (2020–2022) — the nation will face extreme fiscal strain due to import dependency. Forecasts (H4) confirm that the structural deficit will widen even under stable WTI prices, emphasizing the critical need for systemic policy intervention based on non-linear forecasting architectures.

In this context, the strategic evolution of Indonesia's downstream sector necessitates a definitive paradigm shift from conventional management to data-driven governance. As evidenced by the persistent refining capacity deficit and the natural decay of domestic production, the integration of advanced computational frameworks is no longer an optional technical enhancement but a fundamental requirement for national energy security. By leveraging the LSTM architecture, this study demonstrates that high-fidelity forecasting can effectively navigate the complexities of global price volatility and operational constraints, providing a robust foundation for fiscal stability and resource optimization.

Furthermore, this technological transition aligns with the broader regional vision of a resilient and sustainable energy future in Southeast Asia. The implementation of sophisticated predictive models serves as a critical enabler for the IOG 4.0 strategic roadmap, ensuring that policies such as the B40 mandate and Enhanced Oil Recovery (EOR) initiatives are executed with precision. Ultimately, the synergy between strategic digital transformation and holistic policy frameworks will empower Indonesia to mitigate macroeconomic vulnerabilities, foster energy equity, and secure its sovereignty within the increasingly dynamic global energy landscape.

Policy Implications

The validated high-precision forecast serves as a strategic instrument for enhancing Indonesian economic governance in three critical areas:

Strategic Fiscal Planning and Resource Allocation

The primary implication involves optimizing the national budget for energy-related expenditures, as resource procurement is inherently sensitive to global market fluctuations. The government could utilize LSTM-derived insights for proactive budget recalibration and real-time allocation. This predictive sensitivity enables the development of dynamic, risk-adjusted budgeting frameworks. By identifying projected volatility cycles, fiscal authorities can implement strategic hedging measures to stabilize import costs, thereby ensuring price stability and minimizing socio-economic friction during market adjustments.

Enhancing Energy Equity and Domestic Productivity

Addressing the structural supply-demand gap requires high-fidelity forecasting to improve the efficiency of current fiscal interventions. As empirical studies suggest a need for more precise targeting in resource distribution, accurate deficit modeling supports a long-term transition toward direct social assistance and strategic fiscal incentives. This data-driven realignment is essential for fostering domestic production and ensuring the long-term sustainability of the national energy transition, ultimately mitigating broader macroeconomic vulnerabilities.

Accelerating the Digital Roadmap (IOG 4.0)

The performance of the LSTM model directly aligns with the IOG 4.0 strategic framework and the National AI Strategy. Key stakeholders in the energy sector should prioritize integrating advanced architectures (such as LSTM or GRU) into operational planning. Beyond demand forecasting, applying these models to analyze field-level output trends is vital for optimizing Enhanced Oil Recovery (EOR) initiatives and streamlining capital expenditure sequencing to maintain national energy security.

Optimized Implementation of the B40 Mandate

The model's capacity to simulate the impact of the B40 mandate enables evidence-based policy timing. Implementation should be segmented according to regional data profiles. By forecasting localized supply-demand dynamics, the government can prioritize regions with stable consumption patterns for early adoption. This approach ensures that the bio-fuel blending policy acts as a catalyst for sustainability without introducing unintended supply-chain stress in vulnerable areas.

Appendix A: Supplementary Data and Model Details

Table A1

Historical Time-Series Data, 2015–2024

Year	Crude Oil Production (Lifting) (MBPD)	Total Oil Consumption (MBPD)	Net Balance (Deficit) (MBPD)	Annual Average WTI Price ($)
2015	779	1.564	–785	48.07
2016	829	1.572	–743	43.14
2017	804	1.565	–761	51.38
2018	**778**	**1.618**	–840	65.86
2019	**746**	**1.582**	–836	56.90
2020	707	1.400	–693	39.16
2021	660	1.461	–801	68.13
2022	612	1.597	–985	94.90
2023	605	1.604	–999	77.58
2024	580 (Est.)	1.600 (Est.)	–1.020 (Est.)	76.60 (Fct.)

Notes: **MBPD** *denotes Thousand Barrels per Day.* Production data for the 2015–2019 period reflect **actual realized output** to ensure historical accuracy. WTI price indices (2015–2019) represent the **arithmetic annual mean** derived from monthly spot price datasets (EIA) to maintain methodological consistency across the time-series. Figures for 2024 are based on mid-year projections and preliminary industrial reports.
Source: Compiled and processed by W. Muljono, P.A. Setyanto (2026) based on data from the Ministry of Energy and Mineral Resources (ESDM), SKK Migas, and the U.S. Energy Information Administration (EIA).

Table A2

Descriptive Statistics of Input Time Series Data

Metric	Crude Oil Production (Lifting)	Oil Consumption (Total)	Annual Average WTI Price ($ approx.)
Count	5.00	5.00	5.00
Mean	632.88	1532.40	71.27
Std. Dev	50.54	95.52	20.42
Min	580.00	1400.00	39.16
25%	605.40	1461.00	68.13
50% (Median)	612.00	1597.00	76.60
75%	660.00	1600.00	77.58
Max	707.00	1604.00	94.90

Source: W. Muljono, P.A. Setyanto' calculation (2026) based on historical data from the Ministry of Energy and Mineral Resources (ESDM), Republic of Indonesia, and the U.S. Energy Information Administration (EIA), 2020–2024.

Table A3

Optimized LSTM Model Architecture and Hyperparameters

Component	Parameter	Value	Rationale
Input Feature Set (X)	Variables	Crude Oil Production, Total Oil Consumption, Annual Average WTI Price	Three distinct time-series inputs were used to predict the target variable.
Output Target (Y)	Variable	Net Balance (Deficit) (MBPD)	The model is trained to predict the key policy variable: the supply-demand deficit.
Look-back Window (T)	Time Steps	3 Years	A window size of *T*=3 was used, meaning the model predicts the deficit for year *t* using data from years *t*–3, *t*-2, and *t*-1.
Input Shape	(Time Steps, Features)	(3, 3)	Defined by the chosen look-back window and the number of input variables.
Model Type	Structure	Stacked LSTM Network	Two sequential LSTM layers were stacked to enhance the model's capacity to learn complex hierarchical features.
Layer 1 (Input)	Units	100	Contains 100 memory units to process the initial sequence input.
Layer 2 (Hidden)	Units	50	Contains 50 memory units and is essential for extracting deeper temporal patterns.
Output Layer	Type/Units	Dense (1)	A fully connected layer with a single neuron to produce the final single-step forecast (the Deficit for year *t*).
Activation Function	Hidden Layers/Output	ReLU / Linear	Rectified Linear Unit (ReLU) for hidden layers, and Linear for the output layer, as the target is a continuous numerical value (regression).
Optimizer	Algorithm	Adam	The Adaptive Moment Estimation (Adam) optimizer was selected for its efficiency in handling large parameter spaces and converging quickly.
Loss Function	Metric	Mean Squared Error (MSE)	Standard loss function for regression tasks, prioritizing the minimization of large prediction errors.
Training	Epochs / Batch Size	200 / 8	The model was trained for 200 epochs with a small batch size of 8, suitable for the limited size of the time-series dataset

Source: Compiled and optimized by W. Muljono, P.A. Setyanto (2026) through iterative experimental validation and hyperparameter tuning.

Table A4

Exogenous Input Variables for Long-Term Forecasting, 2025–2029

Year	Crude Oil Production (Lifting)	Annual Average WTI Price ($ approx.)
2025	570.0	75.0
2026	560.0	78.0
2027	550.0	76.0
2028	540.0	80.0
2029	530.0	77.0

Source: W. Muljono, P.A. Setyanto' strategic assumptions (2026) derived from historical decay rates and global energy market outlooks from the U.S. Energy Information Administration and the Ministry of Energy and Mineral Resources of the Republic of Indonesia.

Table A5

Key Parameters of the ARIMA Benchmark Model

Term	Scientific Description	Role in the Model
y_t	The value of the time series at the current time period ***t***.	The **dependent variable** being forecasted or modeled.
θ_0	The **intercept term** or **drift parameter**.	Represents the mean value of the series when all other terms are zero.
$\varphi_1, \dots, \varphi_p$	The **Autoregressive (AR) coefficients**.	Parameters defining the impact of the **past values of the series itself.**
$y_{t-1}, \dots, y_{t-p}$	Past observations of the series up to ***p* lags**.	The **Autoregressive component** of order *p*, indicating dependence on *p* previous observations.
ε_t	The **white noise error term** at time *t*.	Represents the **random shock or innovation** at the current time; typically assumed to be independent and identically distributed (i.i.d.) with zero mean and constant variance ($\varepsilon t \sim N(0, \sigma_2)$).
$\theta_1, \dots, \theta_q$	The **Moving Average (MA) coefficients**.	Parameters defining the impact of the past error terms.
$\varepsilon_{t-1}, \dots, \varepsilon_{t-q}$	Past **error terms (shocks)** up to ***q* lags**	The **Moving Average component** of order *q*, indicating dependence on *q* previous error terms

Notes: The ARIMA model combines Autoregressive (AR), Integrated (I, referring to differencing for stationarity), and Moving Average (MA) components. The term *yt* typically represents the differenced series in a non-stationary model. The complete ARIMA model is generally denoted as ARIMA(p,d,q), where *p* is the AR order, *d* is the degree of differencing, and *q* is the MA order.
Source: Adapted from the Box-Jenkins methodology for time-series analysis (Box et al., 2015) and standard econometric validation frameworks.

Table A6

Performance Evaluation Metrics for Model Validation

Metric	Definition and Significance	Ideal Threshold	Comparative Performance Significance
MAPE	Mean Absolute Percentage Error: measures average relative prediction variance.	Minimize	Indicates a 57,24% improvement in forecasting precision over baseline models.
RMSE	Root Mean Square Error: assesses error magnitude in Thousand Barrels Per Day (MBPD).	Minimize	Demonstrates superior handling of non-linear outliers in supply-demand gaps.
R^2	Coefficient of Determination: quantifies the goodness-of-fit on a scale of 0 to 1	Proximity to 1	Yields a high explanatory power (0.96), validating the model's structural fit

Source: Compiled by W. Muljono, P.A. Setyanto (2026) based on experimental validation results.

Table A7

Comparison of Model Performance Metrics

Model Category	Model Name	Application to Indonesian Oil Sector	Key Forecasting Strength
Deep Learning	Long Short-Term Memory (LSTM)	Forecasting volatile Oil Demand and the Deficit Gap, incorporating complex non-linear inputs like GDP, prices, and policy changes.	Captures long-term temporal dependencies and non-linear patterns in time-series data.
Statistical	ARIMA	Modeling the linear portion of the time series, often used to forecast Oil Supply or as a baseline model.	Excellent at capturing linear trends, seasonality, and short-term dependencies.
Hybrid/ Ensemble	ARIMA-LSTM Hybrid	Forecasting the final Deficit Gap by leveraging the strengths of both models to minimize error and increase robustness.	Combines linear and non-linear modeling to provide superior accuracy over single-model approaches.
General ML	Support Vector Regression (SVR)	Forecasting Oil Supply or demand when the available historical data points are limited or if high-dimensional feature selection is required.	Highly effective with sparse or limited data; finds the optimal hyperplane to maximize predictive generalization.
Ensemble (Tree-based)	XGBoost / Random Forest	Feature selection and determining the relative importance of macro-economic indicators (e.g., exchange rates) on Oil Demand	Superior at handling non-linear feature relationships and preventing overfitting through ensemble learning

Notes: This table compares various modeling techniques used to forecast key variables in Indonesia's oil market, specifically focusing on Supply, Demand, and the resulting Deficit Gap. Deep Learning models like LSTM excel with complexity, while Statistical models like ARIMA provide a robust linear foundation. Hybrid and Ensemble methods aim to leverage the strengths of multiple models for improved predictive performance.
Source: (Arab, Benitez, 2025).

References

Aditya, I.A., Wijayanto, T., & Hakam, D.F. (2025). Advancing renewable energy in Indonesia: A comprehensive analysis of challenges, opportunities, and strategic solutions. *Sustainability*, *17*(5), 2216. https://doi.org/10.3390/su17052216 EDN: OODZQX

Ahmad, T., Khan, M.J., & Javed, I. (2022). Artificial intelligence applications in the oil and gas industry: A review. *Journal of Petroleum Science and Engineering*, *212*, 110292. https://doi.org/10.1016/j.petrol.2022.110292

Arab, I., & Benitez, R. (2025). Benchmarking time series forecasting models: From statistical techniques to foundation models in real-world applications [Preprint]. https://doi.org/10.48550/arXiv.2502.03395

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., & Ljung, G.M. (2015). *Time series analysis: Forecasting and control*. Wiley.

Dave, E., Leonardo, A., Jeanice, M., & Hanafiah, N. (2021). Forecasting Indonesia exports using a hybrid model ARIMA-LSTM. *Procedia Computer Science*, *179*, 480–487. https://doi.org/10.1016/j.procs.2021.01.031

Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management*, *10*, 1–9. https://doi.org/10.1177/1847979018808673

Fathin, M.R., Widhiyasana, Y., & Syakrani, N. (2021). Model for predicting electrical energy consumption using ARIMA method. *Advances in Engineering Research*, *207*, 298–303.. https://doi.org/10.2991/aer.k.211106.047

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Jo, Y., Sorkhabi, R.B., & Panja, P. (2026). Machine learning for forecasting production in tight oil reservoirs: Application to the Bakken formation of Williston basin. *Unconventional Resources*, 6, Article 100335. https://doi.org/10.1016/j.uncres.2026.100335

Hanan, A., Swastika, M.B., & Rohmawati, H.M. (2024). *Effective downstream policy strategies for oil, gas, and mining sub-sectors in Indonesia* (Policy Brief No. 5). Purnomo Yusgiantoro Center. https://doi.org/10.33116/pyc-br-5

Hermansah, M., & Canas Rodrigues, P. (2024). Indonesian inflation forecasting with recurrent neural network long short-term memory (RNN-LSTM). *Enthusiastic: International Journal of Applied Statistics and Data Science*, *4*(2), 132–142. https://doi.org/10.20885/enthusiastic.vol4.iss2.art5

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://www.cri.minesparis.psl.eu/people/silber/cours/2025/nlp/bib/HochreiterS1997.pdf

Izudin, N.E.M., Sokkalingam, R., Daud, H., Mardesci, H., & Husin, A. (2021). Forecasting electricity consumption in Malaysia by hybrid ARIMA-ANN. In S.A. Abdul Karim et al. (Eds.), *Proceedings of the 6th International Conference on Fundamental and Applied Sciences* (pp. 749–760). Springer. https://doi.org/10.1007/978-981-16-4513-6_66

Kumoro, M.N.P., Ainin, A.Q., Addinie, I.A., & Nengkoda, A. (2022). A strategy to achieve 1 MMbopd national oil production target by implementation of artificial intelligence algorithm to accelerate the exploration process. In *Proceedings of the Indonesian Petroleum Association, 46th Annual Convention & Exhibition 2022*. Indonesian Petroleum Association. https://doi.org/10.29118/IPA22-SE-157

Lai, G., Chang, W.-C., Yang, Y., & Liu, H. (2018). Modeling long- and short-term temporal patterns with deep neural networks. In *Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval* (pp. 95–104). https://doi.org/10.1145/3209978.3210006

Mayasari, F., & Dalimi, R. (2019). Fuel oil supply demand projection and planning in Indonesia using system dynamics modeling. *International Journal of Smart Grid and Clean Energy*, *8*(1), 11–21. https://doi.org/10.12720/sgce.8.1.11-21

Mukhacheva, A.V. (2025). Econometric modeling of digital development of the economy and social sphere of Russian regions. *π-Economy*, *18*(4), 7–31. (In Russ.). https://doi.org/10.18721/JE.18721/JE.18401 EDN: GYXBSF

Muljono, W., & Setiyawati, S. (2022). Digital economy: The main power for digital industry in Indonesia. *International Journal of Technological Globalisation*, *15*(4), 423–444. https://doi.org/10.1504/IJTGM.2022.125908 EDN: NYMZBZ

Nikitins, I. (2022). Innovative approaches to measure the business environment uncertainty and tools to combat it. *Russian Journal of Innovation Economics*, *12*(1), 223–236. (In Russ.). https://doi.org/10.18334/vinec.12.1.114193 EDN: CWSFHI

Odimarha, A.C., Ayodeji, S.A., & Abaku, E.A. (2024). Machine learning's influence on supply chain and logistics optimization in the oil and gas sector: A comprehensive analysis. *Computer Science & IT Research Journal*, *5*(3), 725–740. https://doi.org/10.51594/csitrj.v5i3.976

Patria, N., Irawanto, B., & Abrar, A.N. (2024). "(Don't) stop the rising oil price": Mediatization, digital discourse, and fuel price controversies in Indonesian online media. *Journalism and Media*, *6*(3), 124. https://doi.org/10.3390/journalmedia6030124 EDN: GLAPIA

Resosudarmo, B.P. (2011). Energy security in Indonesia. In L. Anceschi & J. Symons (Eds.), *Energy security in the era of climate change* (pp. 161–179). Palgrave Macmillan. https://doi.org/10.1057/9780230355361.0019

Ribeiro, A.M.N.C., do Carmo, P.R.X., Rodrigues, I.R., Sadok, D., Lynn, T., & Endo, P.T. (2020). Short-term firm-level energy-consumption forecasting for energy-intensive manufacturing: A comparison of machine learning and deep learning models. *Algorithms*, *13*(11), Article 274. https://doi.org/10.3390/a13110274

Rizvi, M.F. (2024). ARIMA model time series forecasting. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, *12*(5), 3782–3785. https://doi.org/10.22214/ijraset.2024.62416 EDN: UMYIEF

Safrina, R., Kilmi, A., Pradiva, Z.A., & Sohirin, A. (2025). *Enabling a resilient and sustainable ASEAN energy future through digitalisation* (Policy Brief No. 12). ASEAN Centre for Energy (ACE). https://aseanenergy.org/enabling-a-resilient-and-sustainable-asean-energy-future-through-digitalisation/

Sahara, Dermawan, A., Amaliah, S., Irawan, T., & Dilla, S. (2022). Economic impacts of biodiesel policy in Indonesia: A computable general equilibrium approach. *Journal of Economic Structures*, *11*(1), 22. https://doi.org/10.1186/s40008-022-00281-9 EDN: RZPQPO

Setiabudi, P. (2026, January 10). *B40 biodiesel consumption exceeds target in 2025, cuts diesel imports*. Petromindo. https://www.petromindo.com/news/article /b40-biodiesel-consumption-exceeds-target-in-2025-cuts-diesel-imports

Smagulova, K., & James, A.P. (2019). A survey on LSTM memristive neural network architectures and applications. *The European Physical Journal Special Topics*, *228*, 2313–2324. https://doi.org/10.1140/epjst/e2019-900046-x EDN: HRKIKB

Syarovy, M., Nugroho, A.P., Sutiarso, L., Suwardi, Muna, M.S., Wiratmoko, A., Sukarman, & Primananda, S. (2023). Prediction of oil palm production using recurrent neural network long short-term memory (RNN-LSTM). In J. Sumantyo et al. (Eds.), *Proceedings of the International Conference on Smart and Innovative Agriculture (ICoSIA 2022)* (pp. 55–66). Atlantis Press (Springer Nature). 10.2991/978-94-6463-122-7_6

Wibowo, A., Astuti, W., Wibowo, T.P., & Purnomo, H. (2022). Estimation of energy demand in Indonesia using artificial neural network. *International Journal of Energy Economics and Policy*, *12*(6), 261–271. https://doi.org/10.32479/ijeep.11390 EDN: ZFVHWU

Wijaya, I.G.P.O.I., Kristi, M.S., Maharani, N.S., & Raharjo, J. (2025). Data adjustment for ARIMA method to investigate the impact of pandemic on electricity consumption prediction in Indonesia. *Emitor: Jurnal Teknik Elektro*, *25*(1), 47–53. https://doi.org/10.23917/emitor.v25i1.8676

Zega, A.S.R., Hidayat, A.K., Jannah, N.T., & Kartiasih, F. (2024). Selecting the best model for forecasting Indonesia's oil and gas import value using ARIMAX and ARIMAX-LSTM. *Dynamic Management Journal*, *8*(4), 912–941. https://doi.org/10.31000/dmj.v8i4 EDN: DLITGE

Zeinula, A., Khan, Q., Kalam, S., & Riazi, M. (2025). *Hybrid machine learning for oil production forecasting using random forest and long short-term memory models* [Paper presentation]. SPE Annual Caspian Technical Conference and Exhibition, Baku, Azerbaijan. https://doi.org/10.2118/230427-MS

Zhuang, X.Y., Wang, W.D., Su, Y.L., Dai, Z.X., & Yan, B.C. (2025). Deep learning-assisted optimization for enhanced oil recovery and CO2 sequestration considering gas channeling constraints. *Petroleum Science*, 22(8), 3397–3417. https://doi.org/10.1016/j.petsci.2025.04.028

Bio notes / Сведения об авторах

Wiryanta Mujono, PhD, senior researcher in digital economics, Ministry of Communication and Digital of the Republic of Indonesia, 9 Jl. Medan Merdeka Barat, RT.2/RW.3, Gambir, Jakarta Pusat, Jakarta, 10110, Indonesia; associate professor of energy economics, Universitas Sebelas Maret, 36 Jl.Ir. Sutami, Kentingan, Kec. Jebres, Kota Surakarta, Jawa Tengah, 57126, Indonesia. ORCID: 0000-0001-8679-7767. ResearcherID: AAC-4800-2021. E-mail: wiryantamuljono@gmail.com

Padmanabha Adyaksa Setyanto, student, School of Electrical Engineering and Informatics (Computing), Institut Teknologi Bandung, 10 Jl. Ganesa, Lb. Siliwangi, Kecamatan Coblong, Kota Bandung, Jawa Barat, 40132, Indonesia. ORCID: 0009-0004-4639-0961. E-mail: padmanabhaadyaksasetyanto@gmail.com

Муджоно Вирьянта, доктор философии, старший научный сотрудник в области цифровой экономики, Министерство связи и цифровых технологий Республики Индонезия, Индонезия, 1110, Джакарта, ул. Мердека Барат, д. 9; доцент кафедры экономики энергетики, Университет Себелас Марет, Индонезия, 57126, Суракарта, ул. Сутами, д. 36. ORCID: 0000-0001-8679-7767. ResearcherID: AAC-4800-2021. E-mail: wiryantamuljono@gmail.com

Сетьянто Падманабха Адьякша, студент Школы электротехники и информатики (Вычислительной техники), факультета электротехники и информатики (вычислительная техника), Технологический институт Бандунга, Индонезия, 40132, Бандунг, ул. Ганеса, д. 10, Индонезия. ORCID: 0009-0004-4639-0961. E-mail: padmanabhaadyaksasetyanto@gmail.com