INKUBIS: Jurnal Ekonomi dan Bisnis
Volume 8, Issue 2, 1017-1033
e_ISSN: 2775-3913
https://inkubis.poltekssci.ac.id/index.php/ink/index
DOI: doi.org/10.59261/inkubis.v8i2.316

A Machine Learning-Based Early Warning System for Predicting Rupiah Exchange-Rate Pressure in Indonesia

***Welly Ramantika**
Universitas Bina Nusantara
Jakarta,
Indonesia

Muhammad Zarlis
Universitas Bina Nusantara
Jakarta,
Indonesia

***Corresponding author:**
Welly Ramantika, Universitas Bina Nusantara, Indonesia.
✉welly.ramantika@binus.ac.id

Article Info:
***Article history*:**
Received: May 20, 2026
Revised: July 01, 2026
Accepted: July 06, 2026

***Keywords*:**
Early Warning System; Indonesian Economy; Mixed-Frequency Data; Twin Deficit; SHAP; XGBoost

Abstract

Background: Predicting economic vulnerability within the *Twin Deficit Hypothesis* (TDH) framework remains a significant challenge for developing countries due to the frequency mismatch between slow-moving macroeconomic fundamentals and rapidly changing financial market indicators.
Objective: This study aims to address this gap by developing a hybrid *Early Warning System* (EWS) that integrates multi-frequency data to predict the probability of a *Twin Deficit* crisis in Indonesia with a one-month forecasting horizon.
Methods: The dataset comprises macroeconomic and market data covering the period from 2010 to 2026. A mixed-frequency data alignment approach is employed to integrate quarterly fiscal and macroeconomic indicators, monthly fiscal deficit data, and daily market indicators, including the USD/IDR exchange rate, government bond yields, and global commodity prices.
Results: Historical fiscal data gaps (2010–2017) are addressed using ARIMA-based stochastic interpolation. The XGBoost model, optimized using *Borderline-SMOTE*, achieves an accuracy of 79.17%, a recall rate of 87.50%, and a PR-AUC value of 0.9146 in a 24-month out-of-sample evaluation, outperforming *Logit* and *Probit* benchmark models. SHAP analysis identifies global Brent crude oil prices and domestic bond market stress as the dominant factors influencing the occurrence of *Twin Deficit* crises in Indonesia.
Conclusion: These findings demonstrate that integrating high-frequency market signals enhances the detection of emerging fundamental vulnerabilities, providing an operational tool for fiscal and monetary authorities to support preventive policy coordination. The proposed machine-learning-based EWS, which combines mixed-frequency data integration with explainable artificial intelligence (XAI), offers a theoretically grounded and practically applicable framework for macroeconomic crisis monitoring in developing economies. This approach enables forward-looking and coordinated policy responses by Bank Indonesia and the Ministry of Finance of the Republic of Indonesia before fiscal and external imbalances escalate into a full-scale crisis.

To cite this article: Welly Ramantika, Muhammad Zarlis. (2026). A Machine Learning-Based Early Warning System for Predicting Rupiah Exchange-Rate Pressure in Indonesia. *INKUBIS: Jurnal Ekonomi dan Bisnis*, 8 (2), 1017-1033. https://doi.org/10.59261/inkubis.vxix.316

INTRODUCTION

History demonstrates that currency exchange-rate crises have posed a significant threat to national economic stability, often resulting in severe economic downturns, hyperinflation, and diminished public confidence in economic regulators (Gayathri et al., 2024; Salavrakos & Palmadessa, 2023). Indonesia has experienced substantial pressure on the Rupiah exchange rate, which has contributed to periods of economic instability. The Asian Financial Crisis of 1997–1998

and the Global Financial Crisis of 2008 are prominent examples. These events highlight the importance of establishing a robust and timely Early Warning System (EWS) to assist policymakers in anticipating and mitigating exchange-rate fluctuations while maintaining macroeconomic stability.

Traditionally, EWS frameworks have relied heavily on conventional macroeconomic indicators, including foreign exchange reserves and inflation rates (Ferdous et al., 2022). Although these approaches have demonstrated partial effectiveness, most existing studies depend on low-frequency data (monthly or quarterly), which may fail to capture the rapid, high-frequency dynamics commonly observed in currency markets. Furthermore, traditional data mining approaches often experience limited performance when applied to imbalanced datasets because of their difficulty in handling skewed class distributions (Han et al., 2005; Wongvorachan et al., 2023; W. Chen et al., 2024).

This study introduces a novel methodology based on the Twin Deficit Hypothesis (TDH), which proposes a causal relationship between fiscal deficits and current account deficits, emphasizing the importance of accurately identifying minority-class events (i.e., crisis episodes). In open economies that depend heavily on natural resource exports, fiscal deficits may exacerbate current account deficits, while external shocks affecting the current account balance may reduce government revenues and widen fiscal imbalances (Manasseh et al., 2025; Thompson et al., 2021). The TDH tends to demonstrate stronger explanatory power in environments where inflation exceeds 2%, as elevated inflation can contribute to current account deterioration through rising production costs, higher prices of goods and services, and increased import demand (Brixova et al., 2025; Köse & Ünal, 2024). Furthermore, Lutfiansyah (2025) identified a significant negative relationship between inflation and exchange rates in ASEAN countries, indicating that rising domestic inflation contributes to currency depreciation. This study assumes that persistent fiscal deficits may intensify inflationary pressures, which subsequently contribute to exchange-rate depreciation.

Empirical evidence regarding the TDH in developing economies remains inconclusive, highlighting the necessity of country-specific modeling approaches. Hussain (2024) confirmed the validity of the TDH in Pakistan using an Autoregressive Distributed Lag (ARDL) framework, identifying unidirectional causality from fiscal deficits to current account deficits. In contrast, Lubna (2024) identified bidirectional causality in Bangladesh, suggesting that external deficits may also contribute to fiscal imbalances. Similarly, Manasseh (2025) demonstrated that fiscal deficits generate asymmetric effects on current account balances depending on a country's level of commodity dependence. These contrasting findings emphasize the importance of context-specific approaches, particularly for commodity-exporting economies such as Indonesia, where global commodity price fluctuations may directly influence fiscal and external sector linkages. This study extends TDH analysis by applying a machine learning classification framework that represents twin deficit episodes as binary crisis signals, thereby enabling probabilistic, high-frequency crisis prediction rather than relying solely on causal estimation.

Despite increasing interest in EWS development, considerable inconsistencies remain within previous research findings. Studies employing conventional econometric approaches, such as (2022) and Thompson (2021), provide mixed evidence regarding the predictive capability of macroeconomic indicators for currency crises, particularly in developing economies. Meanwhile, machine learning approaches reviewed by Reimann (2024) demonstrate stronger discriminatory performance compared with logistic regression benchmarks but remain relatively underexplored within the twin deficit literature. Moreover, previous studies differ substantially in defining crisis events: some employ threshold-based exchange-rate depreciation measures, whereas others emphasize foreign reserve depletion, making comparisons across studies challenging. The lack of a unified, high-frequency, multi-dimensional crisis definition capable of capturing fiscal–external sector interactions represents a critical research gap addressed by this study.

The research gap within the existing EWS literature can be identified across three dimensions. First, from a methodological perspective, previous studies such as Ferdous (2022) primarily rely on low-frequency monthly or quarterly data, limiting their ability to capture rapid intramonth financial market movements. Second, from a theoretical perspective, the utilization of TDH as a binary classification target for machine learning models remains largely unexplored, particularly in ASEAN economies. Third, from a technical perspective, previous studies have not

simultaneously integrated mixed-frequency data alignment, Borderline-SMOTE oversampling, and SHAP-based model interpretation to construct a transparent and operationally deployable EWS for twin deficit crises. This study addresses these limitations by proposing an integrated analytical pipeline that harmonizes multi-frequency data, manages severe class imbalance, and provides explainable predictions to support monetary and fiscal policy coordination.

This study addresses two primary challenges: (1) examining the applicability of TDH within a high-frequency forecasting environment and determining whether low-frequency macroeconomic indicators can serve as precursors of daily exchange-rate movements; and (2) resolving methodological challenges associated with integrating heterogeneous data frequencies, including the combination of monthly fiscal indicators and quarterly current account data with daily exchange-rate observations. To overcome these challenges, this study applies a mixed-frequency data alignment approach based on linear interpolation, conceptually related to the MIDAS (Mixed Data Sampling) framework, combined with machine learning techniques. MIDAS enables models to incorporate high-frequency information directly without requiring manual aggregation, such as converting monthly observations into quarterly intervals (Moreira, 2025; Shen et al., 2024). This study employs the eXtreme Gradient Boosting (XGBoost) algorithm because of its ability to capture complex nonlinear relationships among variables. To improve model interpretability, SHapley Additive Explanations (SHAP) are applied to quantify the contribution of individual features. SHAP enhances transparency and interpretability in complex machine learning models (Barnes & Hutson, 2024; Deng et al., 2024).

Despite the expanding literature on macroeconomic crisis EWSs, significant challenges remain regarding the integration of heterogeneous data frequencies and the development of crisis definitions that adequately represent structural economic vulnerabilities. Therefore, this study seeks to answer the following research question: How can a monthly Early Warning System that integrates quarterly macroeconomic indicators, monthly fiscal data, and daily financial market information through the proposed mixed-frequency data alignment approach improve the accuracy, timeliness, and interpretability of Twin Deficit crisis prediction as an indicator of Rupiah exchange-rate pressure in Indonesia. More specifically, this study investigates the effectiveness of the proposed mixed-frequency data alignment approach in harmonizing signals across different temporal frequencies, evaluates whether an XGBoost model enhanced with Borderline-SMOTE provides superior predictive performance compared with conventional econometric approaches, and identifies key determinants of crisis occurrence through SHAP-based model interpretation.

Accordingly, this study aims to develop a machine learning-based Early Warning System capable of predicting Twin Deficit crises in Indonesia by constructing a binary crisis indicator, harmonizing quarterly, monthly, and daily datasets through the proposed mixed-frequency data alignment approach, and developing an XGBoost model optimized with Borderline-SMOTE to address severe class imbalance. The model is evaluated using out-of-sample testing over the most recent 24 months based on Accuracy, Recall, F1-score, ROC-AUC, and PR-AUC metrics, while SHAP is employed to improve model transparency by explaining the contribution of each predictor. Academically, this study contributes to the literature by extending the application of the Twin Deficit Hypothesis within an explainable machine learning framework that incorporates mixed-frequency data structures. Practically, the proposed EWS provides a transparent and operational decision-support tool that can strengthen fiscal and monetary policy coordination by enabling earlier detection of macroeconomic vulnerabilities in Indonesia.

The novelty of this study lies in integrating the Twin Deficit Hypothesis with an explainable machine learning framework that combines mixed-frequency data alignment, Borderline-SMOTE, XGBoost, and SHAP to develop a monthly EWS for predicting Rupiah exchange-rate pressure. Unlike previous studies that generally rely on single-frequency datasets or conventional econometric models, the proposed framework integrates quarterly macroeconomic indicators, monthly fiscal variables, and daily financial market information into a unified prediction architecture. This approach enhances predictive performance and model interpretability, providing a transparent and operational framework for macroeconomic crisis monitoring and policy support in developing economies.

The proposed monthly EWS is supported by near-real-time market data, providing monetary policymakers with timely signals to facilitate effective market interventions while enabling fiscal policymakers to monitor the effects of fiscal sustainability on external

vulnerabilities. Therefore, the proposed EWS can serve as an integrated decision-support instrument for monetary and fiscal policy coordination, contributing to more targeted policy responses and strengthening economic resilience.

METHOD

Research Procedures and Techniques

This study adhered to a systematic computational approach designed to handle multi-frequency data and imbalanced class distributions. The procedures were conducted as follows:

Historical data were collected from publicly available sources, including quarterly macroeconomic indicators (Current Account Deficit [CAD], Gross Domestic Product [GDP], and foreign exchange reserves), monthly fiscal balances, and daily financial market variables (JISDOR rates, SBN yields, net foreign flows, U.S. Treasury yields, Brent crude oil prices, and the DXY Index).

A frequency alignment framework was applied to integrate datasets with different temporal frequencies. Interpolation methods (linear or spline interpolation) were used to transform quarterly data into monthly observations. Meanwhile, daily market data were converted into monthly statistical features, including the mean, standard deviation, and momentum, using a 28-day look-back window (tail_n = 28).

The formula for transforming quarterly data into monthly observations was defined as follows:

$$y = y_0 + (x - x_0)\frac{y_1 - y_0}{x_1 - x_0}$$

Where y is the monthly value sought, y_0 and y_1 are the known quarterly values at time points x_0 and x_1.

Daily to monthly data transformation formula (standard deviation):

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

s: sample standard deviation.
x_i: i-th data value
$\bar{x}$: mean sample.
n: sample size.

Setting the dependent variable as a binary classification outcome based on the TDH, where a "crisis" (1) occurred when both the budget balance and the current account balance (CAD) were in deficit. All other conditions were classified as "normal" (0).

Constructing predictive leads and lags, where features from month t were used to predict the crisis condition in month t+1 (lead_months = 1). This process also included the construction of momentum variables, such as Year-over-Year (YoY) growth in foreign exchange reserves and moving averages of market prices.

Employing the Borderline-SMOTE algorithm within the training pipeline to address the limited occurrence of crisis events. This method generated synthetic observations for the minority class (crisis) near the decision boundary, thereby reducing model bias toward normal economic periods. Unlike conventional SMOTE, which oversampled all minority-class observations, Borderline-SMOTE selectively enhanced minority instances located near the classification boundary. These borderline instances were more likely to be misclassified, making them more informative for the classification process (Hairani et al., 2024; Han et al., 2005). The Borderline-SMOTE formulation was defined as follows:

$$s = x_i + \lambda(x_{zi} - x_i), \lambda \sim U(0,1)$$

x_i: native minority sample.
x_{zi}: one of the minority neighbors.
λ: random number between 0 and 1

The selection of XGBoost over alternative machine learning algorithms, such as Random Forest, LightGBM, CatBoost, or Long Short-Term Memory (LSTM) networks, was based on four methodological considerations. First, XGBoost was specifically optimized for structured tabular data, which characterized all variables in this study, whereas LSTM and neural network models generally require substantially larger sample sizes to achieve reliable convergence. Second,

XGBoost's level-wise tree growth provided more stable performance for small-to-medium-sized datasets (below 10,000 observations) compared with LightGBM's leaf-wise growth strategy, which is more susceptible to overfitting in smaller samples (T. Chen & Guestrin, 2016). Third, XGBoost natively supported the scale_pos_weight parameter for class imbalance correction and integrated effectively with external oversampling techniques, such as Borderline-SMOTE, making it suitable for the imbalanced crisis classification task. Fourth, XGBoost's prediction mechanism was fully compatible with SHAP decomposition, enabling theoretically consistent additive feature attribution that supported economic interpretation, which was a central requirement of this study's policy objectives. These considerations collectively justified the selection of XGBoost as the primary predictive engine for the proposed EWS.

The monthly datasets were prepared for XGBoost model training. The model's hyperparameters, including tree depth, learning rate, and the number of trees, were tuned to obtain the optimal configuration. Model performance was evaluated using an out-of-sample test period covering the most recent 24 months, with Accuracy, Recall, F1-score, ROC-AUC, and PR-AUC employed as evaluation metrics. The model outputs were interpreted using SHAP values. This step quantified the contribution of each feature to the final prediction, thereby improving the economic plausibility and interpretability of the model results.

Research Design

The research design of this study was quantitative, predictive, and explanatory, aiming to analyze economic vulnerabilities in Indonesia. The quantitative approach was applied because this study relied on a multi-frequency numerical dataset that integrated quarterly macroeconomic data, monthly fiscal indicators, and daily financial market data into a structured machine learning pipeline. The predictive approach was employed because the primary objective was to forecast exchange-rate pressure, particularly episodes of Twin Deficit crises, one month ahead (lead_months = 1). These episodes served as structural proxies for exchange-rate pressure because simultaneous deficits in fiscal and current accounts indicated heightened external vulnerability. The explanatory design was implemented through SHAP analysis to examine model transparency and provide deeper insights into the determinants of model predictions, such as the influence of global energy prices (Brent crude oil prices) and stress in the domestic bond market (10-year SBN yields) on economic conditions, particularly exchange-rate pressure. This research design enabled the identification and explanation of the economic linkages underlying financial instability.

The scope of this study involved conducting a quantitative analysis of the Indonesian economy using a multi-frequency dataset. Quarterly macroeconomic data were obtained from SEKI Bank Indonesia, including the current account deficit (CAD), gross domestic product (GDP), and foreign exchange reserves, while exchange-rate data between the Rupiah and the US dollar were obtained from the Jakarta Interbank Spot Dollar Rate (JISDOR). Monthly fiscal data were collected from the Ministry of Finance's monthly budget reports. In addition, global financial indicators included Brent crude oil prices, Indonesian government bond yields, the US Dollar Index, net foreign flows, and US Treasury yields, which were retrieved from various international financial databases. The selected sample period covered 2010–2026 because the available data during this period were considered the most complete and consistent compared with other periods.

Data Description

The study utilized quarterly macroeconomic variables, including the current account balance, GDP, and foreign exchange reserves; monthly fiscal balance data; and daily financial market indicators, including the USD/IDR exchange rate, government bond yields, Brent crude oil prices, net foreign capital flows, and the US Dollar Index. A comprehensive description of each variable, including its measurement unit, data frequency, and expected relationship with the crisis variable, was presented in Table 1.

Table 1. Summary of Variables: Definitions, Units, Frequencies, and Expected Relationships

Variable	Definition	Unit	Frequency	Expected Relationship with Crisis
CAD_Val	Current Account Balance	USD Billion	Quarterly	Negative balance → higher crisis probability
GDP_Val	Gross Domestic Product	IDR Trillion	Quarterly	Denominator for ratio calculation; lower GDP → higher vulnerability
Devisa_Val	Foreign Exchange Reserves	USD Billion	Quarterly	Negative: higher reserves reduce crisis risk
Fiscal_Balance	State Budget Balance (Revenue minus Expenditure)	IDR Trillion	Monthly	Fiscal deficit → higher crisis probability
USDIDR_Spot	USD/IDR Exchange Rate (JISDOR)	IDR per USD	Daily	Higher value (weaker Rupiah) → elevated stress
Yield_SBN_10Y	10-Year Indonesian Government Bond Yield	Percent (%)	Daily	Positive: higher yield → rising borrowing cost, investor risk aversion
Yield_Spread_ID_US	Spread between 10Y Indonesian and US Treasury Bond Yields	Basis Points	Daily	Positive: widening spread signals rising country risk premium
Price_Oil_Brent	Brent Crude Oil Price	USD per Barrel	Daily	Positive: higher price → larger import bill and subsidy burden
Net_Foreign_Flow	Net Foreign Capital Flow in Indonesian Financial Markets	IDR Trillion	Daily	Negative (outflow momentum) → higher crisis risk
DXY_Index	US Dollar Index	Index	Daily	Positive: stronger USD → capital outflow pressure on

	emerging markets

Source: Compiled by authors from Bank Indonesia SEKI, Ministry of Finance, JISDOR, Bloomberg (2026)

Data Processing and Transformation Techniques

A major challenge in modeling long-term macroeconomic time series data for Indonesia was the limited availability of historical data. State budget (*Anggaran Pendapatan dan Belanja Negara* [APBN]) expenditure data were published monthly in recent periods; however, from 2010 to 2017, these data were only available in an aggregated annual format.

To address the absence of monthly fiscal data between 2010 and 2017, this study employed a temporal disaggregation approach. Initially, the Denton–Cholette algorithm was considered; however, due to its comparative advantages in preparing data for machine learning applications, the methodology was replaced with ARIMA-based stochastic interpolation. ARIMA methods provide advantages in capturing the internal dynamics of non-stationary economic time series compared with traditional least-squares approaches, such as the Denton method. Furthermore, the ARIMA-based approach ensured that the estimated monthly fiscal balance remained statistically consistent with the observed annual government budget realization. By modeling the underlying time-series structure, this method generated more realistic month-to-month transitions and prevented artificial fluctuations that could potentially introduce misleading patterns during the XGBoost model training process.

The mixed-frequency data alignment approach was employed to synchronize three distinct data layers, namely quarterly macroeconomic indicators (Current Account Deficit [CAD], Gross Domestic Product [GDP], and foreign exchange reserves), which were converted into monthly frequency using linear or spline interpolation through the interp_series function to match the model horizon. Meanwhile, high-frequency financial variables, including USD/IDR exchange rates, Indonesian Government Securities (*Surat Berharga Negara* [SBN]) yields, and Brent crude oil prices, were transformed from daily to monthly observations using statistical measures, including mean, standard deviation, and momentum indicators. These metrics were calculated over a 28-day rolling window (tail_n = 28) to capture short-term market sentiment dynamics.

Data Analysis Techniques

The mixed-frequency data alignment approach served as the primary methodology for integrating quarterly macroeconomic indicators with daily financial market data, thereby producing a consistent monthly forecasting horizon. Quarterly variables were transformed through linear or spline interpolation to generate monthly estimates. Subsequently, daily financial indicators were aggregated into monthly features by extracting statistical moments, including mean, standard deviation, and momentum. This process enabled the model to capture high-frequency market sentiment while maintaining the structural information embedded in lower-frequency macroeconomic indicators.

Next, the Borderline-SMOTE algorithm was applied during the training stage to address the rarity of crisis events within the historical dataset. This technique generated synthetic samples for crisis observations located near the decision boundary, thereby improving class representation. Through this targeted data-balancing approach, the model was able to better distinguish between periods of economic instability and stability, reducing bias toward the dominant “no-crisis” class and improving the detection of significant economic disruptions.

The XGBoost model was trained using the harmonized dataset to capture nonlinear relationships and complex interactions among fiscal, external, and financial market variables. The model configuration was optimized to predict the probability of a Twin Deficit crisis occurring one month ahead. Model performance was evaluated through out-of-sample testing using a 24-month evaluation window. Considering the imbalanced nature of the dataset, this study employed Precision, Recall (Sensitivity), F1-score, Area Under the Receiver Operating Characteristic Curve (ROC-AUC), and Area Under the Precision–Recall Curve (PR-AUC) as evaluation metrics. A confusion matrix was also utilized to assess the trade-off between missed crisis events (false negatives) and unnecessary warnings (false positives). Finally, model interpretation was

conducted using SHAP to quantify the contribution of each independent variable to the predicted probability of a crisis occurrence.

RESULTS AND DISCUSSION

Results

Data Characteristics and Testing Scenarios

The main analysis of this study is based on the model's predictive performance, as evaluated using test set data from the most recent two-year period, specifically from February 2024 to February 2026. The model is configured with a one-month lead time, meaning that predictor variables from month t are used to estimate the probability of a Twin Deficit crisis occurring in the subsequent month (t+1).

The use of the most recent 24 months of historical data as an out-of-sample test set aims to provide a comprehensive evaluation of the model's ability to adapt to economic fluctuations and market volatility during the latest observation period. Over this two-year testing period, the target variable exhibited an imbalanced distribution, consisting of 16 crisis months characterized by simultaneous fiscal and current account deficits and 8 months representing normal economic conditions.

Model Optimization

Model optimization in this study was conducted through two key stages: (1) optimization of data structure and frequency, and (2) predictive feature engineering combined with hyperparameter tuning.

The effectiveness of the XGBoost algorithm as an EWS depends heavily on the availability of short-term momentum and volatility signals. Initial experiments revealed that aggregating predictor variables from monthly to annual frequency resulted in substantially poorer model performance. This deterioration was indicated by an Area Under the Receiver Operating Characteristic Curve (ROC-AUC) value of 0.50, equivalent to random classification performance, along with sensitivity (Recall) values approaching zero. These findings suggest that annual aggregation reduced critical temporal variations in the data, causing excessive smoothing and limiting the model's ability to identify emerging economic stress signals.

Subsequent experiments demonstrated that optimization was achieved by maintaining the original monthly data frequency. Furthermore, replacing the Denton–Cholette benchmarking approach with ARIMA-based stochastic interpolation for estimating missing monthly fiscal data during the 2010–2017 period resulted in improved preservation of seasonal fiscal patterns and intra-year volatility. This adjustment was essential for improving the consistency and stability of model training.

The model's discriminatory capability was further improved through advanced feature engineering and XGBoost hyperparameter tuning. Model optimization was achieved by incorporating high-frequency signals through a 3-month moving average (3m_avg) and momentum deviation indicators to capture the velocity and persistence of economic changes. These engineered features enabled the model to detect structural economic stress before it became evident in lower-frequency indicators. Hyperparameter tuning was performed through systematic experimentation with key parameters, including the learning rate, maximum tree depth, and subsampling ratio, to maximize overall predictive performance.

The classification decision threshold was retained at the default value of 0.50 because alternative threshold adjustments produced only marginal improvements. The resulting performance metrics, including a Recall value of 87.50% and a PR-AUC value of 0.9146, demonstrate that the model achieved high sensitivity through feature enrichment and parameter optimization rather than threshold manipulation. These results provide evidence that the generated early warning signals are derived from robust statistical patterns and reliable probability estimates, thereby supporting their application in monetary and fiscal policy coordination.

Model Performance Evaluation

The confusion matrix provides a detailed assessment of the model's classification performance during the 24-month testing period. The matrix categorizes prediction outcomes

into true positives, true negatives, false negatives, and false positives, which correspond to successful detections, correct classifications of non-crisis periods, missed crisis events, and false alarms. This evaluation framework enables a comprehensive assessment of the system's reliability and practical applicability in real-world early warning scenarios.

Table 2. Confusion Matrix

		Actual	
		Normal (0)	**Crisis (1)**
Prediction	**Normal (0)**	**5 (TN)** *(Correct Rejection)*	**2 (FN)** *(Missed Detection)*
	Crisis (1)	**3 (FP)** *(False Alarm)*	**14 (TP)** *(Correct Detection)*

Source: Data processing results (2026)

Based on the empirical results, the model achieved an overall accuracy of 79.17%. Across 24 observation points from the testing dataset, the model successfully identified 14 of the 16 actual Twin Deficit crisis events. The Recall (sensitivity) value of 87.50% confirms this result. This high detection rate indicates that the system is highly effective in identifying structural economic stress before an actual crisis occurs. The model incorrectly classified three observations as crises when they should have been classified as normal (false alarms) and classified only two observations as normal when they should have been identified as crises (misses). These results demonstrate the model's high precision, with a value of 82.35%, indicating a substantial probability that underlying economic vulnerabilities existed when warnings were generated. In the context of an Early Warning System (EWS), minimizing misses is crucial because the cost of addressing an undetected crisis far outweighs the operational costs associated with false alarms.

The resulting F1 score of 0.848 confirms that the model achieved a strong balance between precision and recall. These results demonstrate that the integration of multi-frequency data through the alignment approach, combined with the nonlinear processing capabilities of XGBoost, significantly enhanced the accuracy of identifying Twin Deficit occurrences in Indonesia compared with traditional econometric baseline models.

From an economic perspective, the confusion matrix outcomes carry significant policy implications within Indonesia's macroeconomic context. The 14 correctly detected crisis episodes corresponded to periods of simultaneous fiscal deficit widening and current account deterioration, which were often associated with surges in global oil prices that placed pressure on both the state budget, through increased energy subsidy costs, and the trade balance through rising import expenditures. The two missed crises (False Negatives) represented months in which the model predicted normal conditions despite the occurrence of actual Twin Deficit conditions, suggesting that, during these periods, high-frequency market indicators had not yet reflected the structural deterioration observed in quarterly current account data.

The three false alarms (False Positives) indicated that market stress signals, such as elevated government bond yields (*Surat Berharga Negara* [SBN]) or capital outflows, occasionally predicted crisis conditions that did not materialize, possibly because Bank Indonesia's timely Triple Intervention policy absorbed external pressures before they translated into structural Twin Deficit conditions. This pattern suggests that the EWS effectively complemented, rather than replaced, human policy judgment, particularly in scenarios where market stress was mitigated through timely monetary intervention.

To further evaluate the model's performance beyond the static threshold, this study referred to Receiver Operating Characteristic (ROC) and Precision-Recall (PR) curve analyses. These visualizations provided a deeper understanding of the model's discriminatory ability across various probability thresholds. The ROC-AUC value of 0.8359 indicated "good" to "excellent" model performance in forecasting macroeconomic conditions. As shown in Figure 1, the curve remained well above the diagonal reference line, indicating that the XGBoost model had a strong ability to distinguish crisis conditions from normal periods. The very high PR-AUC value of 0.9146, as shown in Figure 2, indicated that the model maintained high precision while successfully capturing a large proportion of actual crisis events. The high ROC-AUC and PR-AUC values confirmed the model's strong predictive capability.

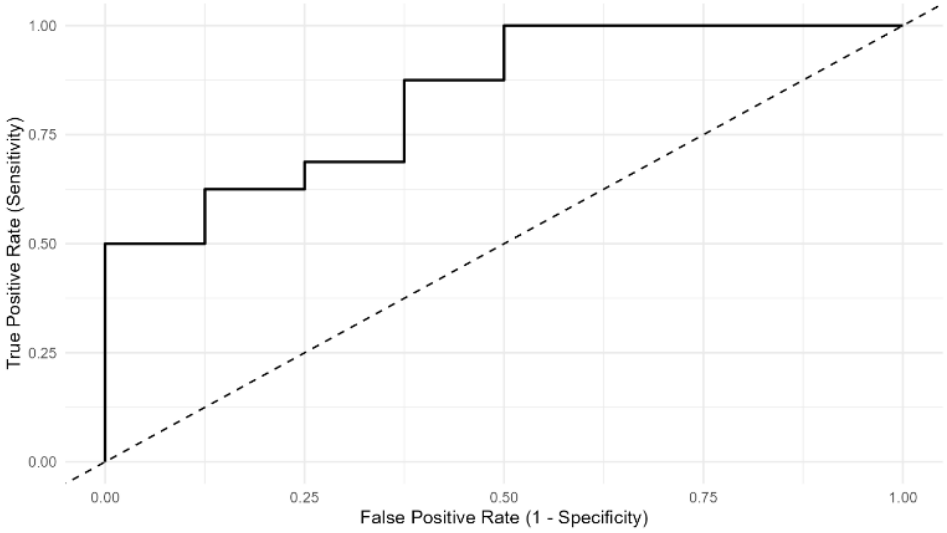

Figure 1. ROC Curve shows that the model has strong discrimination ability.

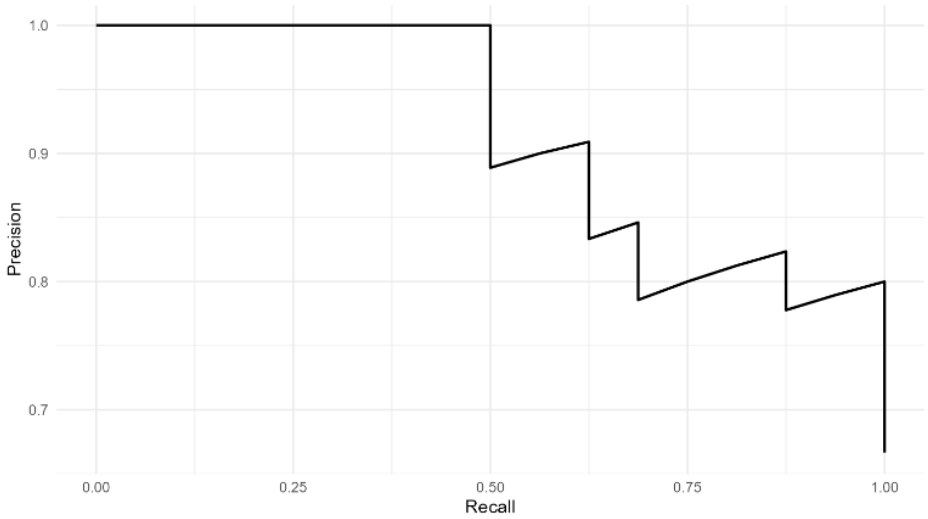

Figure 2. Precision-Recall Curve with a value of 0.9146 shows that the model has high accuracy in detecting twin deficits with few false alarms.

This double-curve validation approach ensures that the EWS is not only accurate in overall prediction performance but also specifically optimized for identifying high-risk economic transitions.

Model Interpretation and Identification of Crisis Determinants (SHAP Analysis)

The application of SHAP to provide transparent and economically grounded interpretations of model predictions ensures that the XGBoost-based EWS does not function as a "black box." Based on the data processing results, a SHAP feature-importance bar chart is generated, as presented in Figure 3.

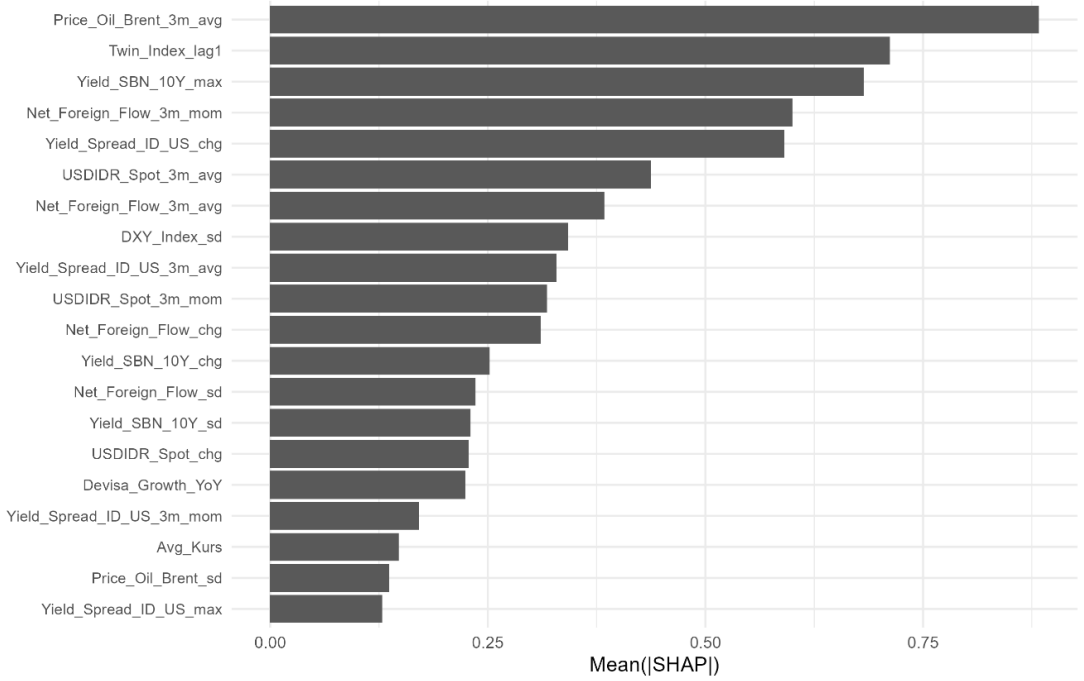

Figure 3. The SHAP Importance bar chart shows that daily market data has a significant impact on the model's predictive ability

The 3-month average Brent oil price is the most significant feature, exhibiting a mean SHAP value of 0.883. This finding is consistent with Indonesia's economic framework, in which variations in global crude oil prices directly affect the fiscal balance through energy subsidies and influence the current account balance through expenditures on oil and gas imports.

The second most influential factor is Twin_Index_lag1, with a mean SHAP value of 0.712. The substantial contribution of this lagged indicator suggests that previous-month crisis conditions are strong predictors of current economic stress.

Yield_SBN_10Y_max is the third key feature, with a mean SHAP value of 0.682. Domestic bond market sentiment plays a crucial role as an early warning indicator. A high 10-year government bond yield signals rising domestic borrowing costs and a potential decline in investor risk tolerance, which may precede an economic contraction.

Net_Foreign_Flow_3m_mom, with a mean SHAP value of 0.599, represents the fourth most influential feature. This metric, which captures momentum in foreign capital flows, serves as an indicator of volatility in investor sentiment. Consequently, abrupt changes in foreign capital movements are strongly associated with external-sector vulnerabilities.

Yield_Spread_ID_US_chg, with a mean SHAP value of 0.591, represents the fifth, but not the final, dominant feature. Changes in the yield spread between Indonesian 10-year government bonds and US Treasury bonds serve as a proxy for the country risk premium. Variations in this spread directly affect capital flows, which subsequently influence the stability of the Rupiah exchange rate.

The influence of external factors, such as oil prices, and market indicators, including SBN yields and capital flows, on domestic macroeconomic conditions highlights the importance of adopting a multi-frequency approach. This finding suggests that, in the short term, market sentiment and global commodity shocks may serve as more immediate drivers of dual-deficit pressures than slower-moving fundamental changes.

The SHAP summary plot provides deeper insights into the direction and magnitude of each variable's contribution to the predicted probability of a Twin Deficit crisis. This visualization enables validation of the underlying economic mechanisms driving the EWS.

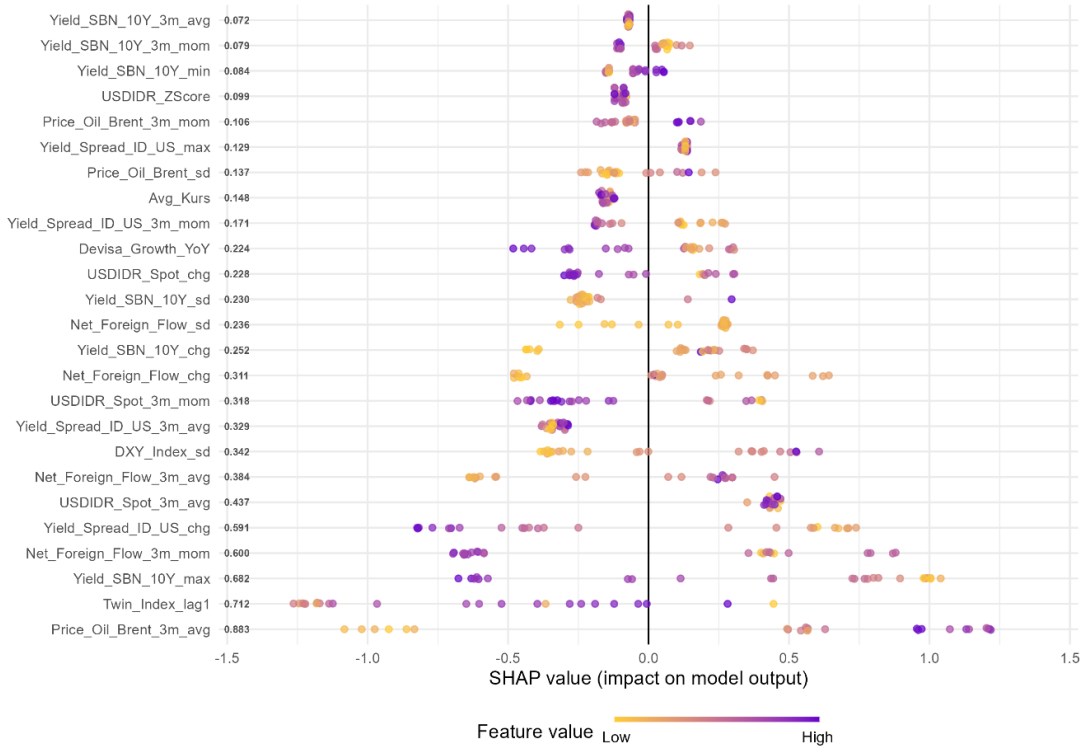

Figure 4. SHAP Summary plot provides deeper analysis than SHAP importance bar chart

Based on the SHAP Summary plot results presented in Figure 4, an analysis of the top predictors was generated, revealing several key economic insights as follows:

Price_Oil_Brent_3m_avg, the most influential feature, indicates that increasing global crude oil prices significantly increases the probability of a Twin Deficit crisis, likely due to the resulting pressure on the oil and gas trade balance and fiscal subsidy expenditures.

The Twin_Index_lag1 variable shows clear clustering, where positive historical index values (deficits) are associated with a higher probability of the current crisis. This finding confirms that structural economic vulnerabilities in Indonesia are likely persistent rather than transient.

High values of Yield_SBN_10Y_max are strongly associated with increased crisis risk. When domestic government bond yields peak, they may trigger heightened investor risk aversion or expectations of rising inflation, thereby increasing the likelihood of entering a Twin Deficit situation.

Negative momentum in Net_Foreign_Flow_3m_mom (outflows) is associated with higher SHAP values. This further demonstrates the role of increasing capital outflows as a precursor to external sector pressures.

The consistency between the SHAP direction and established macroeconomic theory, particularly the Twin Deficit Hypothesis (TDH), validates the model's reliability in predicting Rupiah exchange-rate pressures. The model demonstrates that the constructed Early Warning System (EWS) does not merely memorize historical data but also captures meaningful predictive patterns by assessing the relationships among global commodity prices, market sentiment, and structural fiscal-external sector conditions.

Comparison of the XGBoost model with Logit and Probit

To benchmark the proposed XGBoost-based EWS against conventional econometric approaches, this study compared its predictive performance with Logit and Probit models.

Table 3. Performance Comparison of Logit, Probit, And Xgboost

Metric	Model		
	Logit	***Probit***	***XGBoost***
ROC-AUC	0.8125	0.8281	0.8359
PR-AUC	0.9139	0.9244	0.9146

Accuracy	70.83%	70.83%	79.17%
Precision	0.8	0.8	0.8235
Recall	75%	75%	87.50%
F1	0.7741	0.7741	0.8484

There are differences in the features used among the Logit, Probit, and XGBoost models. XGBoost employs a feature selection procedure to mitigate overfitting. Several significant fundamental macroeconomic variables included in the Logit and Probit models were excluded from the XGBoost model because they were identified as noise, which reduced their contribution to predictive performance. The XGBoost model achieved a higher accuracy rate than the Logit and Probit models despite using fewer variables. This finding indicates that XGBoost is more selective in capturing variables with high predictive power rather than relying solely on statistical significance.

Furthermore, differences exist in feature importance across models in determining predictions. In the Logit and Probit models, predictions are highly dependent on the CAD_Interp value, which represents an interpolation of the current account deficit data published on a quarterly basis. In contrast, XGBoost identifies Brent crude oil prices as the most important feature, and this variable is available at a daily frequency. This finding indicates that XGBoost can capture market signals more rapidly than the Logit and Probit models, making it more responsive to short-term changes in exchange-rate pressure.

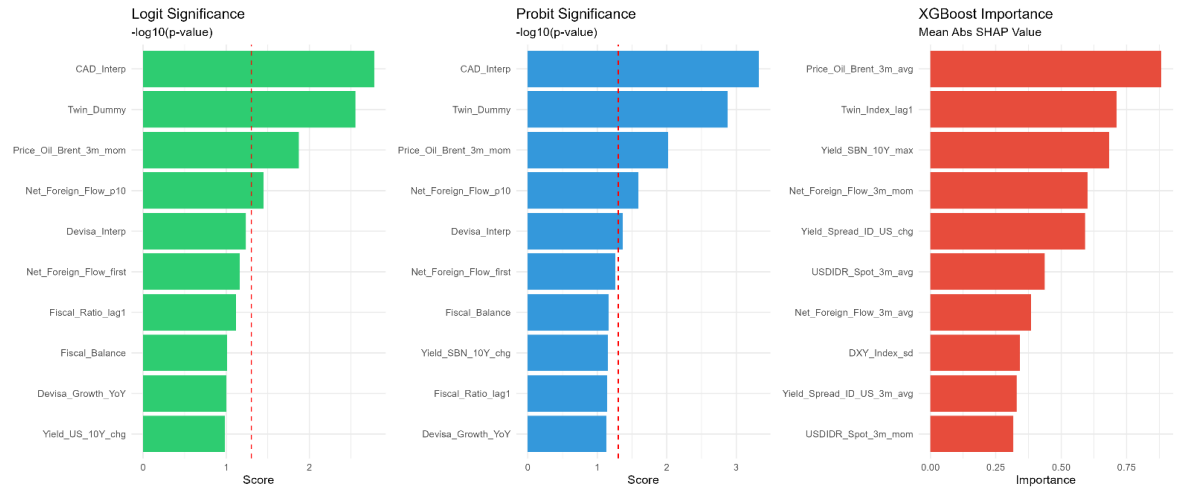

Figure 5. Feature comparison for statistical models (Logit and Probit) vs machine learning models (XGBoost)

While the Logit and Probit models confirm the significance of fundamental variables, such as the current account and fiscal balance, consistent with economic theory, the XGBoost model provides a more relevant representation of current market conditions. XGBoost is able to identify that market-driven variables, such as oil prices and yield spreads, possess stronger predictive power and are more responsive in detecting twin deficit risks than fundamental variables, which are typically published with a time lag.

Discussion

The empirical success of the EWS, characterized by a high recall rate of 87.50% and a PR-AUC value of 0.9146, demonstrates the model's capacity to provide one-month-ahead signals of structural vulnerability, enabling fiscal and monetary authorities to shift from reactive crisis management toward preventive mitigation. Based on the SHAP interpretation results, this study offers three important policy recommendations. (1) The first policy implication is the need to strengthen economic resilience against global commodity price shocks. The dominance of the three-month average Brent oil price indicates Indonesia's continued sensitivity to global energy fluctuations. The Ministry of Finance should utilize EWS signals to assess fiscal sustainability in real time. When rising oil prices increase the probability of a crisis, the government should evaluate the adequacy of energy subsidies and adjust fiscal buffers accordingly to prevent

simultaneous adverse effects on the budget and trade balance. (2) The maximum 10-year SBN yield and the three-month momentum of net foreign capital flows indicate that financial market sentiment provides early signals of deteriorating national economic fundamentals. Bank Indonesia can utilize the EWS as an additional instrument within its "Triple Intervention" strategy. A sharp increase in long-term yields combined with negative capital flow momentum should be interpreted as a leading indicator of structural stress, warranting consideration of targeted market operations to stabilize the bond market before stress escalates into a full-scale twin deficit episode.

According to Lutfiansyah (2025), Bank Indonesia's accommodative interest rate policy and intervention measures have been proven to contribute to exchange-rate stabilization despite challenges related to low foreign exchange reserves and elevated inflation. (3) The success of the mixed-frequency data alignment approach indicates that daily market data contain valuable predictive information for monthly fundamental outcomes. This EWS could be integrated into routine monitoring dashboards used by the Ministry of Finance and Bank Indonesia. Such integration would enable policymakers to implement preventive rather than reactive measures, thereby minimizing the substantial economic recovery costs associated with undetected structural crises and strengthening coordination between monetary and fiscal policies.

The empirical findings align with and extend the theoretical predictions of the Twin Deficit Hypothesis (TDH). The dominance of Brent oil prices and SBN yields as the leading SHAP predictors reflects the transmission mechanism outlined by TDH: rising commodity import costs increase current account pressures while simultaneously placing additional pressure on government subsidy expenditures, thereby widening fiscal deficits. This bidirectional fiscal–external stress mechanism, amplified through the exchange-rate channel, is consistent with the mechanism identified by Manasseh (2025) in commodity-dependent economies.

The strong predictive role of net foreign capital flow momentum corroborates exchange-rate theory, which posits that sudden stops or reversals in capital flows destabilize nominal exchange rates through reductions in foreign currency availability. The application of SHAP values in this study advances the explainable artificial intelligence literature in economics by demonstrating that nonlinear, tree-based models can generate economically coherent feature attributions consistent with established macroeconomic theories, as documented in similar contexts by (Deng, 2024; Reimann, 2024). This convergence between data-driven findings and theoretical predictions strengthens confidence in the model's structural validity beyond mere statistical performance.

A critical practical consideration for EWS implementation is the trade-off between sensitivity (recall) and specificity, which reflects the balance between false alarm rates and missed crisis detection. The current model, with a recall rate of 87.50% and three false alarms over a 24-month period, prioritizes crisis detection at the expense of occasional unnecessary policy alerts. In the context of fiscal and monetary policy coordination, this trade-off is asymmetric: the economic and social costs of an undetected twin deficit episode (including currency depreciation, capital flight, and fiscal consolidation pressures) substantially exceed the operational costs of conducting a precautionary policy review triggered by a false alarm. Bank Indonesia and the Ministry of Finance can mitigate the burden of false alarms by calibrating the decision threshold above 0.50 when fiscal buffers are strong or lowering it during periods of elevated global uncertainty to maximize crisis detection sensitivity. This threshold-adjustment capability, enabled by the model's probabilistic output, provides policymakers with operational flexibility that binary rule-based EWS frameworks cannot offer.

The findings of this study contribute to the rapidly evolving literature on machine learning applications in macroeconomic forecasting. Reimann (2024) demonstrates that ensemble methods, including Random Forest and Extremely Randomized Trees, outperform logistic regression in financial crisis prediction across a broad historical dataset, which is consistent with this study's finding that XGBoost achieves superior accuracy and recall compared with Logit and Probit models. More specifically, the use of SHAP for post-hoc model interpretation aligns with the methodological direction advocated by Deng (2024), who document the utility of interpretable ensemble models in financial risk forecasting. A key distinction of the present study is the integration of mixed-frequency data alignment with machine learning, extending beyond existing EWS frameworks that typically operate on single-frequency datasets. This integration

enables the model to capture both the structural persistence of fiscal–external imbalances and real-time financial market sentiment, representing a methodological advancement over previous approaches.

An important limitation of the proposed EWS is its reliance on the Twin Deficit indicator as the sole proxy for Rupiah exchange-rate pressure. In practice, exchange-rate depreciation may occur independently of twin deficit conditions. For example, sudden geopolitical shocks, global risk-off episodes, or U.S. Federal Reserve policy surprises can trigger substantial Rupiah depreciation without an accompanying deterioration in Indonesia's fiscal or current account positions. This phenomenon was observed during the temporary Rupiah weakness associated with the 2022 global commodity shock and the 2023 U.S. dollar strengthening cycle, both of which affected Indonesian financial markets despite relatively stable fiscal fundamentals. Future research should consider augmenting the crisis definition with additional exchange-rate pressure indicators, such as the Exchange Market Pressure (EMP) index, to capture a broader range of Rupiah vulnerabilities beyond the twin deficit nexus. Such an extension may reduce the risk of structural underprediction in crisis scenarios that are not driven by TDH mechanisms.

CONCLUSION

This study developed and evaluated a machine learning-based Early Warning System (EWS) for identifying Twin Deficit crisis episodes as a proxy for Rupiah exchange-rate pressure in Indonesia. The XGBoost algorithm, optimized using Borderline-SMOTE, demonstrated strong performance in handling class imbalance, achieving 79.17% accuracy and 87.50% recall over a 24-month out-of-sample testing period. The PR-AUC value of 0.9146 further confirmed the model's robust crisis-detection capability with a manageable false alarm rate. SHAP analysis identified global Brent crude oil prices as the dominant predictor of Twin Deficit crises, followed by the 10-year Indonesian Government Securities (*Surat Berharga Negara* [SBN]) yield and net foreign capital flow momentum, supporting the theoretical expectations of the Twin Deficit Hypothesis.

These findings demonstrated that mixed-frequency data alignment combined with explainable machine learning provides a theoretically grounded and operationally feasible framework for macroeconomic crisis early warning in developing economies. This approach extends conventional econometric methods by capturing high-frequency market dynamics that precede structural fiscal and external deterioration. Indonesia's twin deficit structure creates persistent macroeconomic vulnerabilities that are further amplified by daily financial market fluctuations, thereby increasing the risk of Rupiah exchange-rate pressure. The proposed model's one-month-ahead prediction horizon provides Bank Indonesia and the Ministry of Finance with an opportunity to implement proactive and coordinated monetary and fiscal policy responses before structural imbalances escalate into broader economic crises.

Despite its promising predictive performance, this study has several limitations. The proposed Early Warning System relies solely on the Twin Deficit indicator as a proxy for Rupiah exchange-rate pressure, which may not fully capture exchange-rate depreciation arising from geopolitical shocks, global financial market volatility, or unexpected changes in external monetary policy. Future research should therefore improve the model by incorporating additional indicators, such as the Exchange Market Pressure (EMP) index, unstructured data sources including economic news and social media sentiment, and by validating the proposed mixed-frequency XGBoost framework across other developing economies. Furthermore, future studies may explore advanced deep learning architectures, including Long Short-Term Memory (LSTM), Temporal Convolutional Networks (TCN), Transformer-based models, and hybrid ensemble approaches to further enhance predictive performance and model robustness.

ACKNOWLEDGEMENT

The authors would like to express sincere gratitude to Bank Indonesia, the Indonesian Ministry of Finance, and Bloomberg for providing the macroeconomic and financial market data used in this study. The authors also acknowledge the constructive feedback from the anonymous reviewers, which significantly improved the quality of this manuscript.

AUTHOR CONTRIBUTION STATEMENT

Author 1: Conceptualization, Methodology, Software, Formal Analysis, Writing – Original Draft. Author 2: Data Curation, Validation, Writing revisions & Editing, Supervision. All authors have read and agreed to the published version of the manuscript.

REFERENCES

Barnes, E., & Hutson, J. (2024). Navigating the complexities of AI: The critical role of interpretability and explainability in ensuring transparency and trust. *International Journal of Multidisciplinary and Current Educational Research*, *6*(3).

Brixova, B., Šuliková, V., Siničáková, M., & Levandovská, V. (2025). When inflation drives twin deficits catch in the eurozone. *The AMFITEATRU ECONOMIC Journal*, *27*(70), 1127.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chen, W., Yang, K., Yu, Z., Shi, Y., & Chen, C. L. P. (2024). A survey on imbalanced learning: latest research, applications and future directions. *Artificial Intelligence Review*, *57*(6), 137.

Deng, S., Luo, Q., Zhu, Y., Ning, H., & Shimada, T. (2024). Financial risk forewarning with an interpretable ensemble learning approach: An empirical analysis based on Chinese listed companies. *Pacific-Basin Finance Journal*, *85*, 102393.

Ferdous, L. T., Kamal, K. M. M., Ahsan, A., Hong Thuy Hoang, N., & Samaduzzaman, M. (2022). An early warning system for currency crises in emerging countries. *Journal of Risk and Financial Management*, *15*(4), 167.

Gayathri, R., Vijayabanu, C., & Theresa, C. (2024). Economic disruption and global obscurity–Insights and challenges. In *Economic uncertainty in the post-pandemic era* (pp. 1–26). Routledge.

Hairani, H., Widiyaningtyas, T., & Prasetya, D. D. (2024). Addressing class imbalance of health data: a systematic literature review on modified synthetic minority oversampling technique (SMOTE) strategies. *JOIV: International Journal on Informatics Visualization*, *8*(3), 1310–1318. https://doi.org/10.30630/joiv.8.3.2283

Han, H., Wang, W.-Y., & Mao, B.-H. (2005). Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. *International Conference on Intelligent Computing*, 878–887.

Hussain, I., Hayat, U., Alam, M. S., & Khan, U. (2024). A dynamic analysis of the twin-deficit hypothesis: The case of a developing country. *Asia-Pacific Financial Markets*, *31*(1), 25–52. https://doi.org/10.1007/s10690-023-09405-y

Köse, N., & Ünal, E. (2024). The roles of the terms of trade and the real exchange rate in the current account balance. *Economics*, *18*(1), 20220117.

Lubna, M. M., & Saha, S. K. (2024). Justification of the twin deficit hypothesis in Bangladesh. *International Trade, Politics and Development*, *8*(2), 96–116. https://doi.org/10.1108/ITPD-04-2023-0009

Lutfiansyah, W. J., Nikensari, S. I., & Yuniarti, P. (2025). Pengaruh Inflasi dan Suku Bunga Terhadap Nilai Tukar Negara-Negara Regional ASEAN. *AKSIOMA: Jurnal Sains Ekonomi Dan Edukasi*, *2*(8), 2040–2059.

Manasseh, C. O., Logan, C. S., & Ede, K. K. (2025). The impact of budget deficit on current account balance in selected countries. *Journal of Management*, *3*, 233–248.

Moreira, A. (2025). Nowcasting GDP at the Bank of England: a Staggered-Combination MIDAS approach. *Bank of England Macro Technical Paper*, *2*.

Reimann, C. (2024). Predicting financial crises: an evaluation of machine learning algorithms and model explainability for early warning systems. *Review of Evolutionary Political Economy*, *5*(1), 51–83. https://doi.org/10.1007/s43253-024-00114-4

Salavrakos, I.-D., & Palmadessa, A. L. (2023). The global economic crisis: historical roots, lessons learned, and implications for geopolitical stability. In *Globalization, human rights and populism: reimagining people, power and places* (pp. 929–952). Springer.

Shen, F., Yan, X., & Shang, Y. (2024). A novel hybrid PSO-MIDAS model and its application to the US GDP forecast. *PLoS One*, *19*(12), e0315604.

Thompson, K., Sonnylal, L., & Nelson, A. (2021). *The twin deficits hypothesis: A theoretical phantasm or realistic construct*. Research Department Working Paper.

Wongvorachan, T., He, S., & Bulut, O. (2023). A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining. *Information*, *14*(1), 54.