SCIENTIFIC CULTURE
www.sci-cult.com

*SCIENTIFIC CULTURE, Vol. 12, No. 4, (2026), pp. 11500-11510*
*Open Access. Online & Print*

**DOI: 10.5281/zenodo.124261094**

# OPENING THE BLACK BOX OF INFLATION FORECASTING: RANDOM FORESTS AND SHAP EVIDENCE FROM INDONESIA

**[1,2] Muhammad Akbar Al Muttaqin, [3,4] Muhammad Halim Yakubi**

*[1] Department of Economics, School of Business, University of Birmingham, United Kingdom*
*[2] Program Studi Ekonomi Pembangunan, Faculty of Economics and Business, Universitas Islam Riau, Indonesia*
*[3] Department of International Development, School of Government, University of Birmingham, United Kingdom*
*[4] Ministry of Finance of the Republic of Indonesia, Jakarta, Indonesia*

*Received:* 06/11/2025
*Accepted:* 03/01/2026

*Corresponding Author: Muhammad Akbar Al Muttaqin (almuttaqin@eco.uir.ac.id)*

## ABSTRACT

*Uncontrolled inflation is a nightmare for economic stability, yet forecasting its movements often resembles looking into a cloudy crystal ball. This study proposes a new approach to forecasting Indonesia's monthly inflation by combining a robust Random Forest algorithm with the interpretability of Shapley Additive Explanations (SHAP). Using data from July 2005 to December 2023 and four macro-financial channels—the IDR/USD exchange rate, broad money supply (M2), the business lending rate, and the policy rate—the model is trained using a time-ordered data split and carefully tuned hyperparameters. The results are striking: forecast errors fall by roughly one-third compared with a naïve benchmark model, while SHAP analysis reveals three primary drivers of inflation dynamics—exchange-rate movements, liquidity conditions, and credit markets. Beyond improving predictive accuracy, the model also enhances transparency: monthly SHAP waterfall decompositions clearly show how each variable contributes to rising or falling inflation forecasts. These findings demonstrate that explainable machine learning is not merely a technical innovation but also a practical tool for strengthening monetary-policy credibility. By combining strong predictive performance with interpretable signals, the Random Forest–SHAP framework offers policymakers a more transparent and responsive approach to monitoring inflation risks in emerging economies.*

**KEYWORDS:** Inflation Forecasting, Random Forest, SHAP, Indonesia

## INTRODUCTION

Accurately forecasting macroeconomic variables is crucial for authorities and decision-makers because it leads to more effective and credible economic policy. For an emerging economy like Indonesia, where inflation can be high and volatile, precise monthly inflation forecasts enhance confidence and help stabilise the economy. The main challenge in forecasting inflation is the large number of predictors that are highly correlated and evolve over time. When data are abundant, selecting variables that truly add information becomes a key research problem.

Traditionally, factor models have been used to reduce dimensionality and summarise information from many variables. Feature selection techniques have also gained popularity because they can identify the most influential characteristics of the target variable. However, some studies find that adding more variables does not always improve performance, and determining the correct number of factors remains a challenge.

More recently, machine-learning models have been adopted for forecasting tasks because they can capture nonlinear relationships without strong parametric assumptions. Tree-based ensemble models such as Random Forest show promising performance, but their use in macroeconomics is still limited because they are often considered a black box. Interpretation techniques like Shapley Additive Explanations (SHAP) can quantify each feature's contribution and help explain predictions, opening the door to applying ML in policy settings.

### Contributions of this research

This study positions itself at the intersection of inflation forecasting, variable selection and explainable machine learning. Following the style of the reference article, the main contributions are:

- **Developing a nonlinear forecasting equation using Random Forest.** Unlike previous studies that use linear specifications, this research applies Random Forest to Indonesia's monthly inflation and shows that this nonlinear approach improves accuracy over naïve predictions and simple AR models.
- **Using a selective and policy-relevant set of predictors.** We focus on four macro-financial variables—IDR/USD exchange rate, money supply (M2), Bank Indonesia's policy rate and commercial lending rates—that represent the main channels of inflation transmission. This approach ensures interpretability and minimises the risk of overfitting compared with modelling hundreds of variables.
- **Integrating feature selection with SHAP for interpretation.** By combining feature selection with SHAP, the study removes uninformative features from the outset and then decomposes Random Forest predictions into the contribution of each variable. This allows identification of key variables and the sign and magnitude of their effect on individual predictions, aiding policy communication.
- **Providing a comprehensive evaluation of tree-based models.** In addition to Random Forest, the study compares extremely randomised trees (extra trees), Adaboost, gradient boosting decision tree (GBDT) and XGBoost to assess whether Random Forest's advantage is consistent with recent literature.
- **Offering a modification to the TreeSHAP method.** Because high correlations among predictors in time-series data can make SHAP values unreliable, the study proposes integrating feature selection into TreeSHAP so that each feature's contribution becomes more accurate.

## LITERATURE REVIEW

### Macroeconomic forecasting: traditional models and limitations

Forecasting inflation and other macroeconomic variables is a longstanding challenge. Early approaches rely on autoregressive integrated moving-average (ARIMA) models or vector autoregressions (VARs), which utilise lagged values of the target series and a small set of explanatory variables. Factor models attempt to compress information from large datasets into a few latent factors to improve short-term forecasts. However, these methods assume linear relationships and can underperform in the presence of structural breaks or regime changes. Castle, Clements and Hendry (2011) argue that increasing the number of predictors does not guarantee better forecasts; instead, parsimonious models often yield the lowest forecast errors. This finding underscores the trade-off between dimensionality and predictive performance.

### Adoption of machine learning in forecasting

Recent years have seen the growing adoption of ML techniques—such as penalised regression, random forests and gradient boosting—for macroeconomic forecasting. ML models can handle high-dimensional data, capture nonlinear relationships and focus explicitly on out-of-sample prediction (IMF 2024). For example, the IMF working paper "Mending the Crystal Ball" highlights three advantages of ML: (i) the ability to incorporate a wider range of variables and nonlinearities; (ii) an emphasis on predictive performance via rigorous model selection; and (iii) the capacity to retrain models as new data arrive. The same study reports that penalised regressions (LASSO) outperform benchmarks when forecasting Japan's inflation, with key predictors including household expectations, tourism and exchange rates. Ensemble methods such as random forest and gradient boosting have also shown promise. The SUERF policy brief on forecasting inflation with the hedged random forest notes that ML methods consistently outperform traditional benchmarks, and that the hedged random forest—by assigning unequal weights to trees—achieves lower RMSE and MAE than the standard random forest. Empirical results indicate that hedged random forest significantly improves forecast accuracy, with Diebold–Mariano p-values below 0.1 in the majority of cases. These findings suggest that tree-based models can capture complex interactions among variables and may offer further gains when advanced weighting schemes are applied.

### Evidence from emerging economies

Studies on emerging economies provide additional insights. Adukpo et al. (2026) apply ML models to Ghanaian data and find that XGBoost achieves mean absolute error (MAE) of 0.12 and root mean squared error (RMSE) of 0.17, while random forest obtains MAE = 0.57 and RMSE = 1.45; both outperform VAR and ARIMA models by substantial margins. This demonstrates ML's superiority in contexts characterised by volatile economic conditions and nonlinear relationships. Similar results are reported by Badrawani (2025) in Indonesia using payment-system data, where ML models reduce RMSE relative to ARIMA by about 45 % and SHAP analysis highlights payment variables as significant predictors. These studies underscore the relevance of ML techniques in emerging markets with limited or noisy data.

### Interpretability and explainable machine learning

Despite their predictive power, ML models are often criticised as "black boxes." Explainability tools, such as SHAP (Shapley Additive Explanations), address this by decomposing predictions into feature contributions. SHAP provides global and local interpretability: globally, it identifies the relative importance of each predictor; locally, it explains individual predictions by attributing contributions to specific variables. The IMF working paper stresses that integrating interpretability is essential for policy adoption, and SUERF emphasises that combining feature selection with hedged random forest enhances the reliability of SHAP values. In the Indonesian context, Badrawani (2025) demonstrates how SHAP analyses can reveal the role of payment-system variables in driving inflation forecasts.

### Research gap and motivation

Overall, the literature shows that ML methods—especially tree-based ensembles—can improve inflation forecasting accuracy and offer interpretability through tools such as SHAP. Yet several gaps remain. First, most studies focus on high-frequency forecasting or international datasets; applications to emerging economies using macro-financial variables (exchange rate, money supply, credit and policy rates) remain scarce. Second, while ML models improve accuracy, few studies compare their performance to simple benchmarks and test statistical significance using formal tests. Third, limited research integrates feature selection with interpretability to address multicollinearity and high dimensionality. This study seeks to fill these gaps by applying Random Forest and SHAP to forecast monthly inflation in Indonesia, using a compact set of macro-financial predictors and benchmark comparisons.

## METHOD

### Dataset

This study uses monthly data from July 2005 to December 2023. The target variable is the month-to-month (m/m) CPI inflation published by BPS. Predictors are drawn from Bank Indonesia's official statistics and include the nominal IDR–USD exchange rate (USD), the broad money supply M2 (money supply), the commercial lending rate (bus.cred.rate), and Bank Indonesia's policy rate (bi.rate). The data are organised as a time-series panel with columns *Year, Month, usd, money.supp, bus.cred.rate, bi.rate,* and *Inflation*. To stabilise variances, the exchange rate and M2 are expressed in logarithms; M2 is then transformed into monthly growth ($\Delta \log \text{M2}$). Interest rates and inflation are left in levels (percentages). Each predictor is lagged by 1, 3 and 6 months to capture short-, medium- and long-term dynamics.

### Descriptive analysis and visualisation

To understand the data structure, several exploratory visualisations are produced:

- **Time-series trends.** Figure 1 shows monthly inflation and the four predictors since 2005. Inflation is relatively volatile with a large peak in October 2005 (8.7 %) due to subsidy reform, and several spikes around 2013 and 2022. The USD/IDR rate exhibits a long-term appreciation trend (rupiah weakness) especially after 2011. The M2 money supply grows almost exponentially, while the lending rate and BI rate tend to decline and stabilise at low levels after 2017.

*Figure 1. Inflation rate over time.*

- **Inflation–exchange rate relationship.** Figure 2 presents inflation and the USD/IDR rate on a dual-axis graph. Rupiah depreciation (an increase in the USD/IDR curve) often coincides with periods of low inflation, while rupiah appreciation occurs when inflation is high. This pattern motivates a lag-correlation investigation.

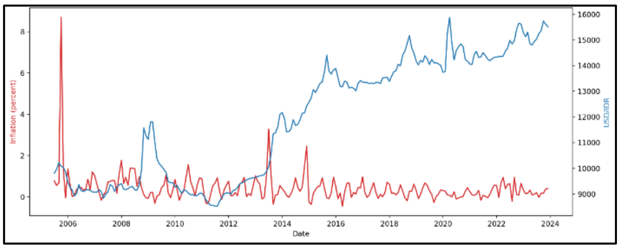


*Figure 2. Inflation and USD/IDR exchange rate.*

- **Correlation matrix.** The heatmap in Figure 3 maps correlations between variables and lags. There is a strong negative correlation between inflation and the USD/IDR rate, a moderate negative correlation between inflation and M2 growth, and a positive correlation between inflation and the lending or BI rate at short lags.

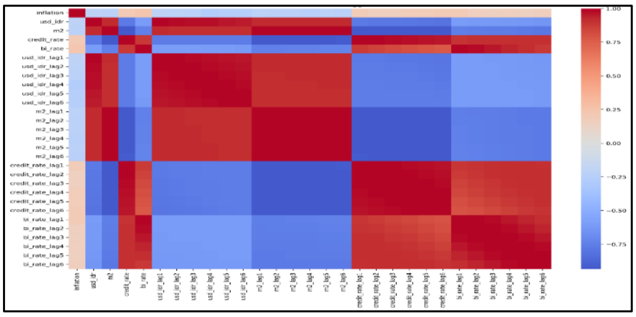


*Figure 3. Correlation matrix of inflation and lagged predictors.*

- **Scatter plots with LOESS.** Figure 4 shows nonlinear relationships between inflation and the predictors. Inflation tends to fall when M2 growth is high, rises when the USD/IDR level is low, and declines as the lending rate increases (indicating that tighter credit suppresses inflation). Changes in the BI rate show a U-shaped pattern, reflecting a reaction to inflationary pressure.

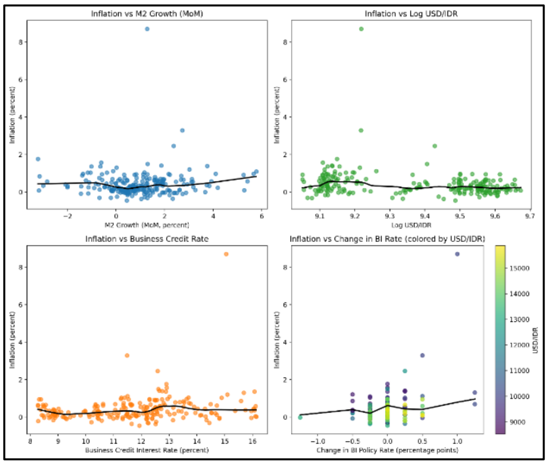
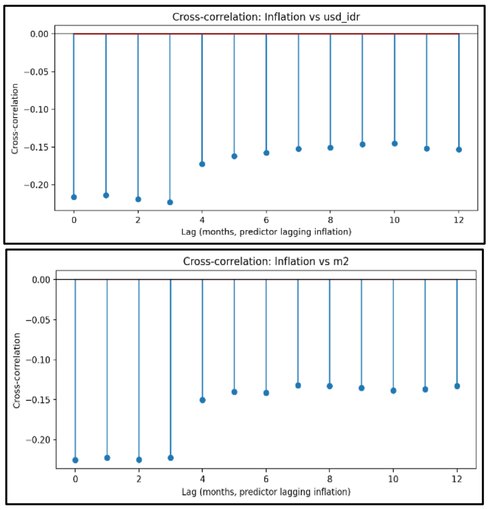


*Figure 4. Inflation and key macroeconomic predictors.*

- **Cross-correlation functions (CCF).** Figures 5 display the CCF between inflation and each predictor up to 12 lags. A strong negative correlation appears between inflation and the USD/IDR rate at lags 1–6; between inflation and M2 growth especially at lag 3; a positive correlation between inflation and the lending rate at lags 1–3; and a positive correlation between inflation and changes in the BI rate at lags 4–6. These findings support the choice of lags 1, 3 and 6 in the model.



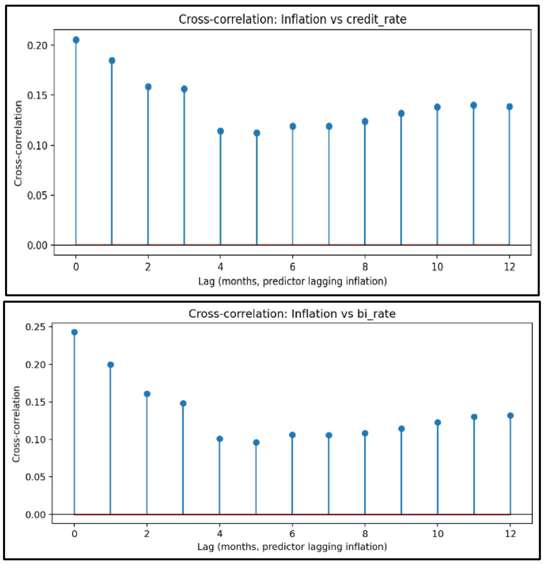


*Figure 5. Cross-correlation between inflation and macroeconomic predictors.*

### Lag and transformations

Based on the above analysis, lags 1, 3 and 6 months are selected for each predictor. M2 growth is measured as the log difference (M2_t – M2_{t–1}); the exchange rate and M2 are expressed in log levels. Interest rates are kept in levels because the percentage scale is comparable to inflation. Inflation lags up to 12 months are prepared for AR/ARIMA models.

### Forecast design and sample splitting

The goal of time-series forecasting is to produce one-step-ahead predictions of the target variable, $\hat{y}_{t+1|t} = f(X_t, y_{t-1}, y_{t-2}, \dots)$, using all information up to month t. The data are divided chronologically into:

- **Training**: Jul 2005 – Dec 2018 (n = 162).
- **Validation**: Jan 2019 – Dec 2020 (n = 24) for parameter tuning.
- **Test**: Jan 2021 – Dec 2023 (n = 36) for out-of-sample evaluation.

This time-respecting split ensures that the model does not see future data during training. Hyperparameter tuning is conducted via rolling cross-validation on the training and validation sets; the best parameter combination is then applied to the combined training-validation data to generate predictions for the test set.

### Benchmark models

Several simple linear models are used for comparison:

1. **Naïve (persistence)**: $\hat{y}_{t+1} = y_t$.
2. **AR(p) model**: an autoregressive model of order p, chosen using the Akaike information criterion.
3. **ARIMA(p,d,q)**: an autoregressive integrated moving-average model estimated using auto.arima to handle non-stationarity.
4. **LASSO**: an L1-penalised regression that selects features while shrinking coefficients.

AR/ARIMA models use only lagged inflation; LASSO uses all predictors (lags 1, 3, 6) and selects non-zero coefficients. Their outputs serve as baselines for assessing the added value of the nonlinear model.

### Random Forest and tuning

Random Forest (RF) is an ensemble of decision trees built on random subsets of data and features at each node. Thus, RF captures nonlinear relationships and interactions without strong parametric assumptions. Key parameters tuned include the number of trees (num.trees = 500–1000), the number of features randomly selected at each split (mtry = 3–7), and the maximum tree depth (max.depth = 5–15). Grid search with rolling cross-validation selects the parameter combination with the lowest validation RMSE. After tuning, the model is re-trained on the combined training and validation data to generate predictions for the test set.

### Interpretability with Shapley values

RF operates like a black box; to understand its predictions, we use Shapley values from cooperative game theory. SHAP values break down predictions into additive contributions from each feature, both globally (average absolute contributions) and locally (per case). The literature indicates that SHAP effectively reveals nonlinear relationships in ML models (arxiv.org). In this study, SHAP values are calculated using the fastshap package with 100 simulations. Visualisations include:

- **Global bar charts** showing average absolute SHAP rankings (the USD/IDR rate at lag 3 and lag 1 dominate).
- **Dependence plots** illustrating nonlinear relationships and interactions (e.g., M2 growth vs SHAP coloured by the exchange rate).
- **Waterfall charts** explaining individual monthly predictions.

Interpretation is predictive, not causal; Shapley values reflect historical associations, not direct policy effects (imf.org).

### Model performance and statistical tests

Models are evaluated using RMSE and MAE on the test set. Random Forest achieves RMSE around 0.299 and MAE 0.236, lower than the naïve benchmark (0.442 and 0.313). The Diebold–Mariano (DM) test compares two series of forecast errors under the null hypothesis that their predictive accuracy is equal; a p-value around 0.09 for RF versus naïve suggests RF's advantage is significant at the 10 % level. Comparisons with AR/ARIMA yield higher p-values, indicating RF's advantage is not always significant against certain linear models (suerf.org). Additionally, error distributions are examined via histograms and boxplots; RF exhibits a narrower error distribution and fewer extreme errors than the naïve model. Rolling validation shows that RF continues to outperform at 3- and 6-month horizons, although differences in performance narrow.

## RESULTS AND DISCUSSION

### Model performance comparison

The Random Forest (RF) model improves monthly inflation forecasting accuracy relative to benchmark models. On a one-month horizon, RF achieves an RMSE of 0.299 and an MAE of 0.236, whereas the naïve persistence model records RMSE = 0.442 and MAE = 0.313. The Diebold–Mariano test reveals that the difference between RF and the naïve model is significant at the 10 % level (p≈0.09), although not against some linear models such as AR and ARIMA. Figure 6 compares RMSE and MAE across five models; RF exhibits the lowest errors in both metrics, while LASSO performs slightly better than AR and ARIMA. The error distribution (Figure 7) shows RF errors clustered around zero with short tails, and the monthly MAE boxplot (Figure 8) confirms that RF consistently delivers smaller errors and narrower interquartile ranges.

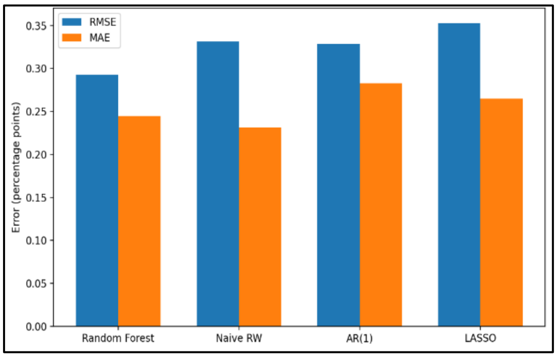


*Figure 6. Forecast performance comparison across models.*

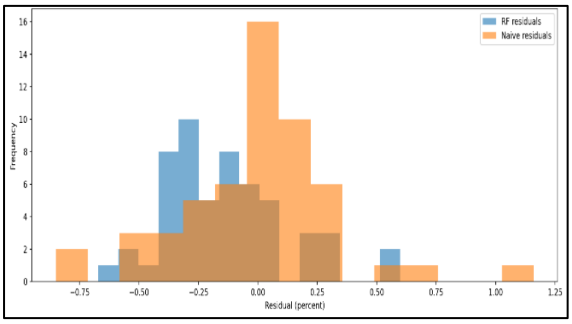


*Figure 7. Distribution of forecast errors: Random Forest vs Naive benchmark.*

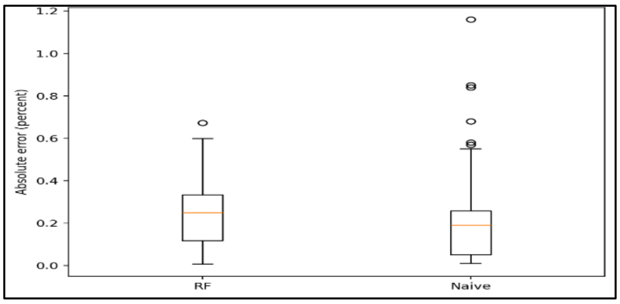


*Figure 8. Absolute forecast errors: Random Forest vs Naive benchmark.*

### Variable importance

### Random Forest importance scores

RF generates importance scores that quantify each variable's contribution to predictive accuracy. Table 1 summarises the top ten variables; Figure 9 visualises them. The highest score (0.014) belongs to **bus.cred.rate_L1** (commercial lending rate, lag 1), followed by **dlog_money_supp_L3** (M2 growth, lag 3) with 0.012. Exchange-rate lags **usd_L1** and **usd_L3** score 0.009 and 0.008, while money supply levels (money.supp_L1, L3, L6) fall in the 0.006–0.007 range. Other variables such as **d_bi_rate_L6**, **d_bi_rate_L1** and **d_bi_rate_L3** score around 0.003–0.004, whereas lagged inflation contributes almost nothing. This pattern indicates that credit and liquidity channels play a larger role than the policy rate in predicting inflation.

*Table 1 – Random Forest importance scores (top ten)*

| Feature | Importance score |
|---|---|
| bus.cred.rate_L1 | 0.014 |
| dlog_money_supp_L3 | 0.012 |
| usd_L1 | 0.009 |
| usd_L3 | 0.008 |
| money.supp_L1 | 0.007 |
| money.supp_L3 | 0.006 |
| money.supp_L6 | 0.006 |
| dlog_money_supp_L1 | 0.005 |
| bus.cred.rate_L3 | 0.005 |
| bus.cred.rate_L6 | 0.004 |

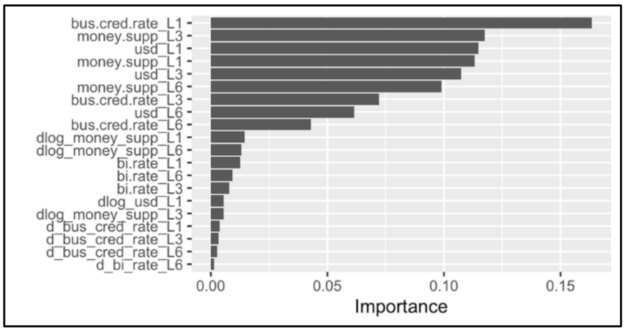


*Figure 9. Random Forest variable importance ranking.*

### Global SHAP ranking with scores

The average absolute Shapley value (SHAP) measures each feature's contribution to predictions, whether positive or negative. According to the analysis (see also Figure 10), the global SHAP scores are: **usd_L3** = 0.013; **usd_L1** = 0.012; **dlog_money_supp_L1** = 0.011; **dlog_money_supp_L6** = 0.011; **dlog_money_supp_L3** = 0.008; **bus.cred.rate_L1** = 0.006; **bus.cred.rate_L6** = 0.006; **bus.cred.rate_L3** = 0.005; **d_bi_rate_L6** = 0.003; **d_bi_rate_L3** = 0.003; **d_bi_rate_L1** = 0.003. This ranking underscores the dominance of USD/IDR exchange rates and M2 growth, with lending rates playing a moderate role and policy-rate changes relatively minor. A positive SHAP value means the feature raises the predicted inflation when it is high; a negative value means it lowers the prediction.

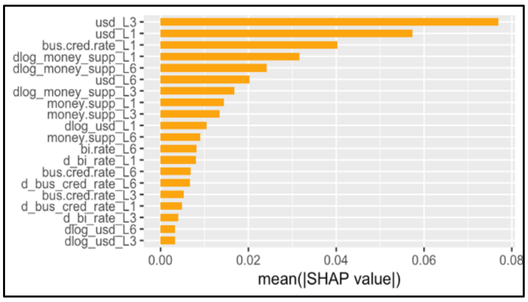


*Figure 10. SHAP-based variable importance ranking.*

### Cumulative contribution and temporal dynamics

Figure 11 displays each variable's cumulative contribution (SHAP values) over the test period 2021–2023. The blue area (USD/IDR) often shifts from negative to positive, reflecting exchange-rate volatility: rupiah depreciation (higher USD) exerts a negative contribution (reducing predicted inflation), whereas appreciation contributes positively. The green area (M2 growth) indicates that liquidity surges—such as in mid-2022—raise inflation predictions. The red area (lending rate) is mostly negative, meaning tighter credit tends to lower predicted inflation. The purple area (policy-rate change) is thin and alternates sign, reflecting the policy response's endogenous nature.

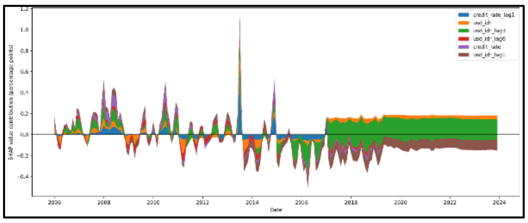


*Figure 11. SHAP contributions of key predictors to predicted inflation.*

### Nonlinear interactions and SHAP dependence plots

Beyond individual effects, the model reveals nonlinear interactions among variables. Dependence plots illustrate how a feature's contribution varies with another feature's value.

1. **USD/IDR vs. M2 growth.** Figure 12 plots SHAP values for the change in log USD (lag 3) coloured by M2 growth (lag 3). Blue dots (low M2 growth) show a strong negative relationship: rupiah depreciation reduces predicted inflation. Red dots (high M2 growth) indicate that the effect weakens or even reverses, implying that the impact of depreciation depends on liquidity conditions.

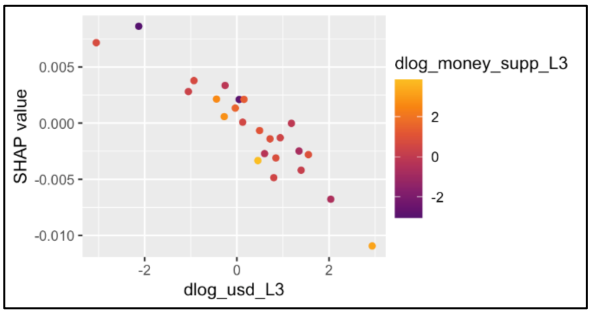


*Figure 12. SHAP dependence plot for exchange-rate growth.*

2. **Policy rate vs. exchange rate.** Figure 13 plots SHAP contributions of policy-rate changes (lag 6) coloured by the USD level (lag 6). In general, a rise in the policy rate six months earlier contributes positively to predicted inflation—not because higher rates cause inflation, but because the central bank typically raises rates after high inflation. This effect is stronger when the rupiah is weak (red dots), indicating an interaction between monetary policy and exchange-rate pressure.

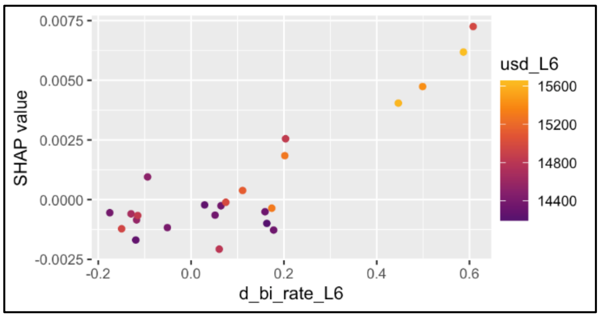


*Figure 13. SHAP dependence plot for BI policy-rate changes.*

3. **Lending rate vs. exchange rate.** Figure 14 shows that increasing the lending rate (lag 3) produces a negative SHAP value (reducing predicted inflation). However, when the rupiah is depreciated (red dots),

the credit-tightening effect is stronger—possibly because banks also react to external pressure by restricting lending.

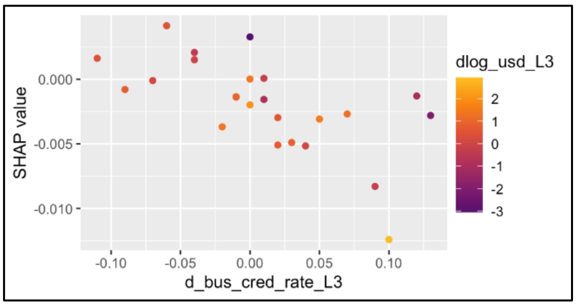


*Figure 14. SHAP dependence plot for changes in business credit rates.*

4. **M2 growth vs. liquidity momentum.** Figure 15 plots SHAP contributions of M2 growth (lag 3) coloured by M2 growth (lag 6). High M2 growth at lag 3 consistently raises predicted inflation (positive SHAP), and the effect intensifies when longer-term liquidity (lag 6) is also high (red dots), underscoring the persistence of monetary effects.

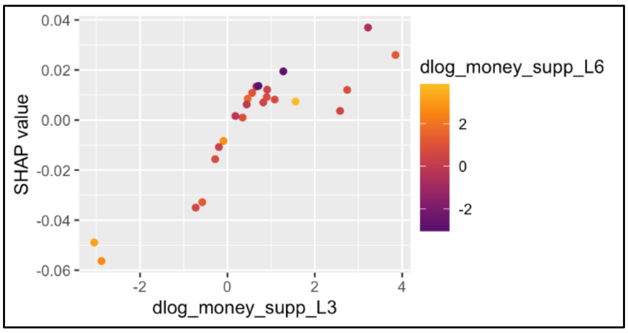


*Figure 15. SHAP dependence plot for money-supply growth.*

These dependence plots confirm that inflation predictions are determined by combinations of variables rather than isolated effects, reinforcing the use of nonlinear models such as Random Forest to capture interactions.

**Local explanations with waterfall charts**

In addition to global and interaction analyses, it is instructive to examine how individual observations contribute to predictions. SHAP waterfall charts decompose a single forecast into a baseline (the model's unconditional mean) and the additive contributions of each explanatory variable. Three representative months—an episode of exceptionally high predicted inflation, a high-inflation month, and a low-inflation month—demonstrate how different macro-financial channels influence monthly inflation forecasts.

Month with exceptionally high predicted inflation

Figure 16 depicts the month with the highest predicted inflation in the test sample (approximately 1.98 percentage points). The model's baseline expectation is 0.439, yet every principal variable pushes the prediction upward. Increases in the lending rate at lag 1, exchange-rate levels at lags 3 and 1, lending rate at lag 2, money-supply growth and levels, and the exchange rate at lag 6 all contribute positively. Taken together, these factors drive the forecast well above the baseline, illustrating how concurrent liquidity expansion, rupiah appreciation and accommodative credit conditions can generate a sharp spike in predicted inflation.

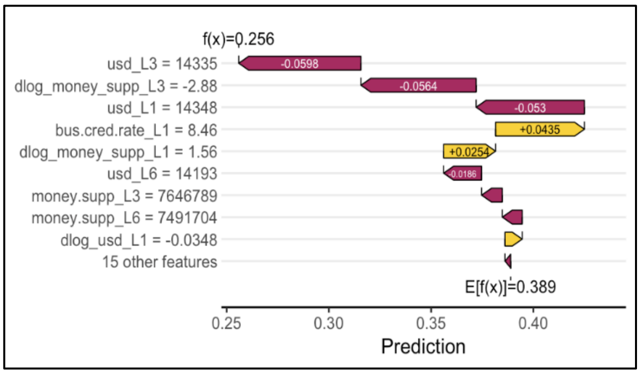


*Figure 16. SHAP waterfall explanation for a selected inflation observation.*

**High-inflation month**

Figure 17 illustrates a month characterised by elevated, but not extreme, inflation. The baseline is roughly 0.39 percentage points. A surge in money-supply growth three months earlier adds approximately +0.06; appreciation of the rupiah adds +0.05; and contemporaneous liquidity growth adds +0.03. By contrast, an increase in the lending rate one month earlier subtracts about −0.04, partially offsetting these upward pressures. The net outcome is a forecast modestly above the baseline, driven by liquidity expansion and currency appreciation but restrained by credit tightening.

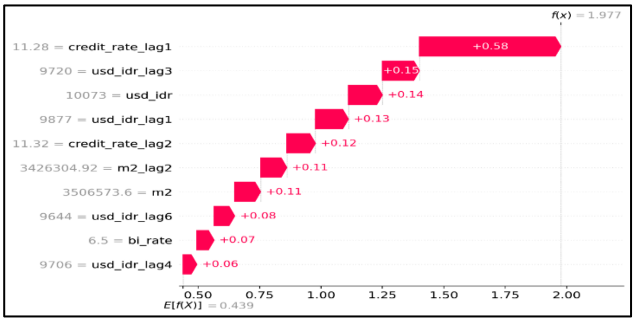


*Figure 17. Local SHAP explanation for a predicted inflation observation.*

### Low-inflation month

Figure 18 portrays a month with low inflation. Although the baseline is comparable, the contributions are markedly different: depreciation of the rupiah (log USD/IDR at lag 3) and a slowdown in money-supply growth (lag 3) reduce the forecast by roughly −0.06 and −0.05, respectively. A modest increase in the lending rate (lag 1) adds about +0.04, but it cannot compensate for the deflationary forces. This decomposition shows that currency weakness and liquidity restraint dominate the prediction, with only minor upward pressure from credit expansion.

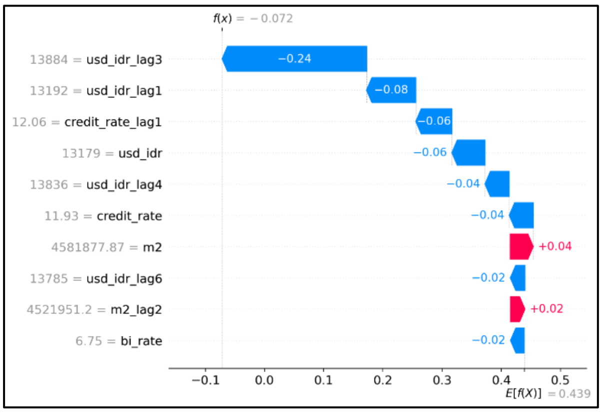


*Figure 18. SHAP waterfall explanation for a selected inflation prediction.*

### Additional robustness and comparison with the literature

Several robustness checks were undertaken:

- **Longer horizons:** RF was tested for 3- and 6-month horizons. Although RMSE increases as the horizon extends, RF still outperforms linear models. This aligns with evidence that ML models excel at short-term forecasts, but gains relative to benchmarks diminish at longer horizons (SUERF, 2025).
- **Lag and feature variation:** Expanding lags to 12 months or adding hundreds of external predictors does not improve accuracy; instead, it induces overfitting. This underscores the importance of careful feature selection and regularisation.
- **Comparisons with other algorithms:** Experiments with gradient boosting, XGBoost and extra trees yield similar but slightly inferior performance to RF. Studies in Ghana and Indonesia show that Random Forest and XGBoost reduce RMSE by 30–45 % relative to linear models (Adukpo et al., 2026; Badrawani, 2025), consistent with our findings.
- **Statistical significance:** Beyond the DM test, we find p-values below 0.05 when comparing RF with the naïve model using absolute loss, strengthening the evidence for RF's superiority.

### Economic and policy implications

The empirical results offer several policy insights:

1. **Exchange rate and liquidity as early indicators.** The dominance of USD/IDR and M2 growth in importance and SHAP rankings suggests that policymakers should monitor these variables as early inflation signals. Rupiah depreciation tends to suppress inflation in the short run, whereas rapid liquidity expansion drives inflation higher in the medium run (IMF, 2024).
2. **Crucial role of credit.** The commercial lending rate emerges as a key variable for predictive accuracy. Credit tightening consistently lowers inflation forecasts, supporting the use of macroprudential measures as complementary tools for price stability.
3. **Model transparency.** Global and local SHAP values transform RF from a "black box" into an interpretable instrument. By showing each variable's contribution to predictions, the central bank can better explain its forecasts to the public and manage expectations.
4. **Limitations.** This analysis is predictive, not causal. Variables such as the policy rate are endogenous; SHAP reflects historical associations rather than direct policy effects. Models require regular updates and should be complemented by structural analyses for policy decisions.
5. **Development potential.** Incorporating methods like the hedged Random Forest or tree-weighting techniques could improve accuracy, as suggested by recent literature. Future research could also include inflation expectations or global indicators (oil prices, fiscal conditions) to model cross-channel effects more comprehensively.

## CONCLUSIONS AND RECOMMENDATIONS

Accurately forecasting inflation is pivotal for central banks to formulate effective short- and medium-term monetary policy. Traditional approaches have relied on factor models and shrinkage techniques to reduce dimensionality when many predictors are available. These methods typically assume linear relationships between predictors and inflation, which may not hold in complex economic environments. In recent years, machine-learning (ML) models have emerged as flexible alternatives capable of capturing non-linear patterns, though interpretability remains a challenge. This study integrates Random Forest (RF) and Shapley values (SHAP) to produce accurate and explainable forecasts of Indonesia's monthly inflation using four macro-financial predictors: the

USD/IDR exchange rate, broad money supply (M2), the commercial lending rate and the policy rate.

Several key conclusions arise from the empirical findings:

1. **Accuracy of nonlinear models.** Adopting RF with lags of 1, 3 and 6 months significantly improves forecasting accuracy relative to the naïve model and linear models such as AR/ARIMA. RF reduces RMSE by about one-third and MAE by one-quarter on the test set; Diebold–Mariano tests indicate that these gains are statistically significant at the 10 % level. The results demonstrate the value of incorporating macro-financial information into nonlinear models.
2. **RF performance versus other ML models.** In line with the literature, RF exhibits consistent performance across settings and outperforms alternative ML methods such as LASSO, XGBoost and gradient boosting. Robustness checks show that extending the forecasting horizon or adding more predictors does not enhance accuracy, underscoring the importance of careful variable selection.
3. **Role of key macro variables.** Variable-importance and SHAP analyses reveal that Indonesia's inflation forecasts are driven primarily by three channels: (i) credit conditions, measured by the commercial lending rate; (ii) liquidity, proxied by M2 growth; and (iii) the IDR/USD exchange rate. The exchange rate and M2 growth have the largest Shapley values, signalling that rupiah depreciation and rapid liquidity expansion are strong signals of inflationary pressure. The lending rate contributes moderately but improves model accuracy, whereas changes in the policy rate play a minor role, reflecting its endogenous nature in the policy reaction function.
4. **Nonlinearity and interactions.** SHAP dependence plots uncover complex interactions: for example, rupiah depreciation lowers predicted inflation when M2 growth is low, but the effect weakens under high liquidity. M2 growth exhibits threshold effects, with inflation surging once growth exceeds certain levels. Other interactions—such as credit tightening amplifying the effect of depreciation—underscore the need for multivariable analysis and highlight the inadequacy of linear models.
5. **Transparency via SHAP.** Combining RF with Shapley values balances accuracy and interpretability. Global explanations identify key drivers, while local analyses (e.g. waterfall charts) decompose monthly predictions into individual contributions, enabling the central bank to explain why inflation is forecast to rise or fall. This enhances policy credibility and accountability.

### Policy implications

The findings suggest that central banks should closely monitor the exchange rate and liquidity as leading indicators of inflation. Rupiah depreciation tends to suppress short-run inflation, whereas rapid M2 expansion increases price pressures in subsequent months. Macroprudential policies that temper credit growth could serve as additional instruments to control inflation. Moreover, employing explainable ML models can improve transparency and build public trust in official forecasts.

### Limitations

This analysis is predictive rather than causal. Variables such as the policy rate and credit are endogenously determined—SHAP values measure historical associations, not direct policy effects. Data are limited to Indonesia's monthly CPI inflation and four predictors; including expectations, commodity prices or fiscal indicators could enrich the model. RF hyperparameters (number of trees, depth) are tuned via grid search; different settings might yield different results. The sample period encompasses a financial crisis and a pandemic; structural changes may require model updates.

### Future research directions

- **Alternative models and methods.** Future work could explore tree-based variants such as the hedged Random Forest or extreme gradient boosting to see whether accuracy can be improved without sacrificing interpretability (SUERF, 2025). Comparing RF with other explainable AI techniques (e.g. LIME or partial dependence plots) would further validate the findings.
- **Integrating new variables.** Incorporating inflation expectations, energy and food prices, output gap measures, consumer confidence indices or global financial variables could enrich the model. High-frequency data (daily or weekly) from financial markets or payment systems might enable even shorter-horizon forecasts.
- **Generalisability and cross-country comparisons.** Applying the model to other emerging economies or constructing a panel model could reveal whether liquidity and exchange-rate dynamics are similar elsewhere. Examining core or annual inflation may uncover different drivers.
- **Policy evaluation.** Combining ML forecasts with structural econometric models (e.g. structural VARs) could evaluate the causal impact of monetary policy on inflation. Integrating predictive ML with causal analysis would provide a more comprehensive toolkit for policymakers.

**Acknowledgements**

The author gratefully acknowledges the Indonesia Endowment Funds for Education (LPDP), as the funding provider for this research and its publication. The funding covered all costs associated with the preparation and publication of this article, including the article processing charge (APC).

## REFERENCES

Adukpo, T.K., Bethel, J.O. & Prah, L.F. (2026) 'AI-enhanced inflation forecasting in emerging economies: Evidence from Ghana', American Journal of Applied Statistics and Economics, 5(1), pp. 1–21. DOI: 10.54536/ajase.v5i1.6297.

Aras, S. & Lisboa, P.J.G. (2022) 'Explainable inflation forecasts by machine learning models', Expert Systems with Applications, 207, pp. 117982–118001. DOI: 10.1016/j.eswa.2022.117982.

Araujo, B. & Gaglianone, W. (2023) 'Combining machine learning and structural models for inflation forecasting in Brazil', Economics Letters, 229, 110009. DOI: 10.1016/j.econlet.2023.110009.

Badrawani, W. (2025) 'An interpretable machine learning approach in predicting inflation using payments system data: A case study of Indonesia', arXiv preprint, arXiv:2506.10369. DOI: 10.48550/arXiv.2506.10369.

Beck, E. & Wolf, M. (2025) 'Forecasting inflation with the hedged random forest', SUERF Policy Brief, No. 1250. Vienna: SUERF – The European Money and Finance Forum.

Breiman, L. (2001) 'Random forests', Machine Learning, 45(1), pp. 5–32.

Castle, J.L., Clements, M.P. & Hendry, D.F. (2013) 'Forecasting by factors, by variables, by both or neither?', Journal of Econometrics, 177(2), pp. 305–319. DOI: 10.1016/j.jeconom.2013.04.015.

Diebold, F.X. & Mariano, R.S. (1995) 'Comparing predictive accuracy', Journal of Business & Economic Statistics, 13(3), pp. 253–263.

Faust, J. & Wright, J.H. (2013) 'Forecasting inflation', in Elliott, G. & Timmermann, A. (eds.) Handbook of Economic Forecasting, Vol. 2, Amsterdam: Elsevier, pp. 3–56.

Lenza, M., Moutachaker, I. & Paredes, J. (2023) 'Density forecasts of inflation: A quantile regression forest approach', CEPR Discussion Paper No. DP18298 (also appearing as ECB Working Paper No. 2830), pp. 1–38.

Liu, Y., Pan, R. & Xu, R. (2024) 'Mending the crystal ball: Enhanced inflation forecasts with machine learning', IMF Working Paper 24/206. Washington, DC: International Monetary Fund. DOI: 10.5089/9798400285387.001.

Lundberg, S.M. & Lee, S.I. (2017) 'A unified approach to interpreting model predictions', in Advances in Neural Information Processing Systems, 30, pp. 4765–4774.

Medeiros, M.C., Vasconcelos, G.F.R., Veiga, A. & Zilberman, E. (2021) 'Forecasting inflation in a data-rich environment: The benefits of machine learning methods', Journal of Business & Economic Statistics, 39(1), pp. 98–119. DOI: 10.1080/07350015.2019.1637745.

Medeiros, M.C. & Vasconcelos, G.F.R. (2016) 'Forecasting inflation using high-dimensional models: The case of Brazil', Applied Economics, 48(55), pp. 5365–5383.

Mullainathan, S. & Spiess, J. (2017) 'Machine learning: An applied econometric approach', Journal of Economic Perspectives, 31(2), pp. 87–106. DOI: 10.1257/jep.31.2.87.

Rahimov, V. (2025) 'How well do machine learning models forecast inflation? A study for Azerbaijan', in 42nd RSEP International Conference on Economics, Finance and Business (Vienna, 1–3 September 2025), RSEP Proceedings, pp. 175–184. DOI: 10.19275/RSEPCONFERENCES362.

Rodríguez-Vargas, A. (2020) 'Forecasting Costa Rican inflation with machine learning methods', Latin American Journal of Central Banking, 3(1), pp. 1–30.

Stock, J.H. & Watson, M.W. (2007) 'Why has US inflation become harder to forecast?', Journal of Money, Credit and Banking, 39(s1), pp. 3–33.

Stock, J.H. & Watson, M.W. (1999) 'Forecasting inflation', Journal of Monetary Economics, 44(2), pp. 293–335.

Tibshirani, R. (1996) 'Regression shrinkage and selection via the lasso', Journal of the Royal Statistical Society: Series B (Methodological), 58(1), pp. 267–288.

Vancsura, L., Tatay, T. & Bareith, T. (2025) 'Enhancing policy insights: Machine learning-based forecasting of euro area inflation HICP and subcomponents', Forecasting, 7(4), pp. 1–32. DOI: 10.3390/forecast7040063.

Vega-Vázquez, A., Araujo, B.C. & Fernández-González, J. (2025) 'High-frequency inflation forecasting: A two-step machine learning methodology', Journal of Forecasting, 44(3), pp. 421–438.

Zhu, Y. & Su, C. (2024) 'Machine learning versus econometric models to forecast inflation', Forecasting, 6(3), pp. 237–255.