Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Evaluation of Policy-Relevant Forecasting Advantage of Machine and Deep Learning Models over Classical Econometrics: Evidence from Ghanaian Headline Inflation

Isaac Nyame
Kwame Nkrumah University of Science and Technology
Baffour Osei
University of Skills Training and Entrepreneurial Development (USTED)
Abigail Boatemaa
University of Bradford
Gabriel Osei Forkuo
gabriel.forkuo@unitbv.ro
Transilvania University of Brasov

Research Article

Keywords: ARIMA, central banking, structural breaks, autoregressive persistence, sub-Saharan Africa, out-of-sample stress-testing

Posted Date: July 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10480821/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

This study evaluates whether machine learning and deep learning models offer a genuine forecasting advantage over classical econometric benchmarks for Ghana's headline (year-on-year) inflation, a question of direct relevance to a central bank operating an inflation-targeting regime in a small, commodity-exposed, currency-volatile economy. Using 280 months (January 2000–April 2023) of Bank of Ghana monetary and macro-financial data, we construct a strictly one-step-ahead forecasting matrix in which every predictor reflects information available one month before the forecast target, and compare nine models (a random walk and seasonal naive baseline, an AIC-selected ARIMA benchmark, two regularized linear models [Ridge, Lasso], two tree ensembles [Random Forest, XGBoost], a kernel method [Support Vector Regression], and a shallow feed-forward neural network) on a chronological train/validation/test split whose 42-month test window (October 2019–April 2023) deliberately spans Ghana's 2022–23 inflation surge. Contrary to the presumption that more flexible models dominate, the ARIMA(3,1,0) benchmark achieved the lowest test-set RMSE (2.41 percentage points), narrowly ahead of the random walk (2.65) and Lasso (3.41). However, Diebold–Mariano tests show that ARIMA is not statistically distinguishable from either of these two closest competitors, while its advantage over every other model is statistically significant. Every nonlinear machine learning model underperformed both simple baselines out of sample, with several models (notably XGBoost and the neural network) showing a large gap between near-perfect training fit and poor test performance: a textbook overfitting signature accentuated by the regime shift embedded in the test period. We argue this is a genuine and policy-relevant finding rather than a failure of specification: headline inflation is highly persistent, and models that best exploit that persistence, rather than those with the greatest functional flexibility, forecast best through a structural break. The study details a leakage-free feature-engineering protocol, reports full hyperparameter-tuning and diagnostic results, and discusses the conditions, such as larger samples, richer nonlinear feature sets, or regime-aware architectures, under which machine learning might be expected to add value for inflation forecasting in a market such as Ghana's.

JEL Classification: C53, C45, E31, E37, E52, O55

1. Introduction

Inflation forecasting occupies a central place in the operational toolkit of any central bank pursuing price stability, and the stakes are particularly high in small, open, commodity-exposed economies such as Ghana, where headline inflation is buffeted simultaneously by exchange-rate pass-through, imported food and energy prices, and domestic monetary conditions (Acquah-Sam, 2017; Ohemeng et al., 2021). Ghana's recent experience makes the point starkly: after several years in single digits, headline year-on-year inflation rose from roughly 10 percent in early 2022 to a peak above 50 percent within a year, driven by currency depreciation, fiscal pressures, and global commodity shocks (Kunawotor et al., 2021; International Monetary Fund, Fiscal Affairs Department, 2018; International Monetary Fund, Monetary and Capital Markets Department, 2020a, 2020b; World Bank Group, 2020). Forecasting through episodes

of this kind, not just during calm, low-inflation periods, is precisely when accurate forecasts matter most for monetary policy committees, yet it is also when forecasting models are most severely tested.

Two broad forecasting traditions are available. The classical econometric tradition builds on autoregressive and structural time-series models, such as ARIMA and SARIMA specifications (Peychinov et al., 2025), vector autoregressions, and error-correction frameworks, that have long been the default tools for inflation forecasting in African central banks and are documented, for example, in SARIMA-based studies of Nigerian inflation (Otu et al., 2014) and Kenyan inflation (Nyamao, 2014), as well as in the IMF's food/non-food inflation forecasting framework for low-income countries (Andrle et al., 2013). The second, more recent tradition draws on machine learning and deep learning, which have been applied with encouraging results to related African macro-financial series (Osei-Dwomo & Forkuo, 2026), including a long short-term memory (LSTM) network for the Ghanaian cedi–US dollar exchange rate (Adekoya et al., 2021) and a stacked-ensemble approach to decomposing inflation drivers in Nigeria (Akande et al., 2021). The premise animating much of this second literature, that greater model flexibility should translate into better predictive accuracy, is intuitively appealing but has not, to our knowledge, been tested rigorously for Ghana's headline inflation using a forecasting design that (a) excludes any contemporaneous or look-ahead information, (b) tunes every model's hyperparameters via walk-forward validation rather than in-sample fit, and (c) evaluates out-of-sample performance across a genuine structural break rather than a calm holdout period chosen for convenience.

This study fills that gap. Using 23 years of monthly Bank of Ghana monetary, interest-rate, and macro-financial data, we build a leakage-free, one-step-ahead forecasting matrix for headline year-on-year inflation and compare nine candidate models spanning naive baselines, a classical ARIMA benchmark, and six machine-learning and deep-learning specifications. Models are tuned exclusively on a training window (2001–2017), compared on a held-out validation window (2017–2019), and evaluated exactly once on a 42-month test window (2019–2023) that was deliberately selected to include Ghana's 2022–23 inflation surge: a deliberately hard, policy-relevant stress test rather than a favorable holdout. We report the full result honestly, including where the more flexible machine-learning models underperform simple benchmarks, and we discuss why this outcome is informative rather than a specification failure.

The remainder of the study is organized as follows. Section 2 reviews the literature on inflation dynamics in Ghana and Sub-Saharan Africa, classical time-series forecasting, and machine-learning applications to macro-financial forecasting. Section 3 describes the data, cleaning procedure, feature engineering, train/validation/test design, candidate models, and evaluation metrics. Section 4 presents descriptive statistics, model performance comparisons, overfitting diagnostics, feature importance, and residual diagnostics. Section 5 discusses the results in light of the literature and their implications for forecasting practice at the Bank of Ghana. Section 6 concludes, states the study's limitations, and outlines directions for future research.

2. Literature Review

2.1 Inflation Dynamics and Monetary Policy in Ghana and Sub-Saharan Africa

Ghana adopted a formal inflation-targeting framework in the 2000s, and its design and early performance have been examined through both Taylor-rule and full-fledged inflation-targeting lenses (Boamah, 2012, as cited in related Ghanaian monetary-policy literature; Clinton et al., 2010). Acquah-Sam (2017) identifies money supply growth, exchange-rate depreciation, and fiscal deficits as the dominant proximate drivers of Ghanaian inflation, a finding echoed in broader West African evidence on exchange-rate pass-through and inflation dynamics (Ohemeng et al., 2021). At the regional level, studies of the West African Monetary Zone document how banking-sector fragility and inflation dynamics interact (Adeleke & Ogebe, as reflected in WAMZ-focused literature), while cross-country evidence from the Economic Community of West African States points to only partial and slow inflation convergence across the zone (Muhammed-Gani et al., 2022), underscoring the country-specific idiosyncrasies that make a Ghana-specific forecasting exercise worthwhile rather than one that can be generalized wholesale from neighboring economies. Kunawotor et al. (2021) further show that extreme weather events have a measurable, policy-relevant effect on African inflation through agricultural-supply channels, a mechanism directly pertinent to Ghana's food-price-sensitive consumer basket.

Beyond Ghana, the broader Sub-Saharan African inflation-targeting experience offers useful comparative context. Studies of Sierra Leone (Jackson et al., 2020), Mauritius (Madhou et al., 2021), Rwanda (Ngarambe et al., 2016; World Bank, 2020b), Nigeria (George-Anokwuru, 2023; World Bank, 2020a), and Uganda (International Monetary Fund, 2013, 2018) collectively suggest that inflation-targeting credibility, exchange-rate management, and fiscal discipline interact in ways that are highly country-specific, reinforcing the case for models tailored to Ghana's own data-generating process rather than imported specifications calibrated elsewhere.

2.2 Classical Time-Series Approaches to Inflation Forecasting

The autoregressive integrated moving average (ARIMA) family remains a workhorse for inflation forecasting in developing economies precisely because headline inflation series are typically highly persistent and close to a unit-root process, conditions under which simple autoregressive structures are difficult to beat. Otu et al. (2014) apply SARIMA models to Nigerian inflation and report strong in-sample and short-horizon forecasting performance; Nyamao (2014) reaches a similar conclusion for Kenyan inflation using a multiplicative ARIMA specification. At the policy-institution level, Andrle et al. (2013) develop a food/non-food decomposition forecasting framework for low-income countries at the IMF, explicitly built around the recognition that African inflation series combine a persistent core component with volatile, supply-shock-driven food and energy components. Charemza et al. (2014) caution more generally that inflation forecast risk is often asymmetric and non-systematic, meaning point-forecast accuracy alone (as emphasized in this study's RMSE/MAE/MAPE metrics) does not capture the full forecasting-risk picture faced by a policy committee: a caveat we return to in the discussion.

2.3 Machine Learning and Deep Learning in Macro-Financial Forecasting

A growing literature applies machine learning and deep learning methods to African macro-financial series. Most directly relevant to this study, Adekoya et al. (2021) train a long short-term memory (LSTM) network to forecast the Ghanaian cedi–US dollar exchange rate and report that the deep-learning model outperforms simpler benchmarks for that series. Akande et al. (2021) use a stacked-ensemble approach, combining several base learners, to decompose and predict the drivers of Nigerian inflation, again reporting favorable results relative to single-model benchmarks. These findings are frequently cited as evidence that machine learning "works" for African macro forecasting; however, both studies target series (an exchange rate; decomposed inflation components in-sample) and employ validation designs that differ materially from the strict, leakage-free, crisis-spanning out-of-sample test constructed in this study for aggregate headline inflation. Tassi (2019) situates this broader shift toward data-rich, algorithmic forecasting within the wider "big data" turn in economic statistics, noting that the availability of granular, high-frequency data is a necessary but not sufficient condition for machine learning to outperform parsimonious classical models; sufficiency also requires a sample size and signal-to-noise ratio adequate to the flexibility of the model being fit, a condition that, as this study's results show, is not automatically satisfied by a 200-month monthly training window.

2.4 Comparative Inflation-Targeting and Monetary-Policy Regimes

A separate strand of literature evaluates inflation-targeting regimes across emerging and developing economies more broadly, providing comparative context for interpreting Ghana's policy environment. Svensson (2010) lays out the canonical full-fledged inflation-targeting framework against which country implementations are typically benchmarked. Chugunov et al. (2019, 2021) examine the macroeconomic effects of inflation targeting across both advanced and developing economies, finding effects that are highly sensitive to institutional credibility and fiscal-monetary coordination. Country-specific studies, South Africa (Ncanywa & Ralarala, 2020; Oladipo, 2017; World Bank, 2012), Brazil (Filho & Milan, 2018), and India (Islam & Ahmed, 2023), similarly find that inflation-targeting effectiveness is conditional on exchange-rate regime, fiscal stance, and central-bank credibility, all factors directly relevant to interpreting Ghana's 2022–23 inflation surge and the difficulty any forecasting model faces in anticipating a shock of that magnitude.

2.5 Positioning of the Present Study

This study contributes to the intersection of these three strands by (i) constructing what is, to our knowledge, the most methodologically rigorous leakage-free one-step-ahead machine-learning forecasting design applied to Ghanaian headline inflation to date; (ii) evaluating nine models, including both classical and deep-learning specifications, on a genuinely adversarial, crisis-spanning test period rather than a favorable calm-period holdout; and (iii) reporting results honestly, including instances

where flexible machine-learning models underperform simple baselines, a finding with direct implications for how the Bank of Ghana and similar institutions should weigh model complexity against robustness when selecting operational forecasting tools.

3. Data and Methodology

3.1 Data Sourcing and Cleaning

Data were sourced from 19 monthly bulletin exports published by the Bank of Ghana (Monthly Interest Rates, Monthly Monetary Survey, and Time Series Monthly series), several of which were overlapping or duplicated downloads of the same underlying series. All files were merged and exact-duplicate rows removed, after which each file was reshaped from a wide year-by-month format into a tidy long format (Date, Variable, Value). A systematic data-quality check revealed that every variable's "0.00" entries fell exclusively in the trailing (most recent) months of that variable's series, consistent with not-yet-reported months rather than genuine zero readings, and these were recoded to missing throughout. One genuine data conflict was identified and resolved: the Monetary Policy Rate for May 2018 appeared as both 17.00 and 18.00 percent across two source files; cross-referencing Bank of Ghana Monetary Policy Committee decision history (an 18 percent rate held from March through April 2018, cut to 17 percent in May 2018) resolved the conflict in favor of 17.00.

From the cleaned panel, 24 core macro-monetary variables (Table 1) with continuous monthly coverage from January 2000 through April 2023 (280 months) were retained, spanning headline inflation, policy and market interest rates, monetary aggregates, private-sector credit, international commodity prices (crude oil, cocoa, gold), and the Bank of Ghana Composite Index of Economic Activity. This 280-month core window was chosen specifically because it is the longest span over which these variables are reported without interruption, giving the machine-learning models a materially larger training sample than the shorter (52-month) fully complete cross-section identified in our companion data-preparation workbook.

Table 1
The 24 Core Macro-Monetary Variables, January 2000–April 2023

Category	Variables included
Inflation	Headline Inflation (YoY %) [target variable]
Interest Rates	Monetary Policy Rate; 91- and 182-Day Treasury Bill discount and interest-rate-equivalent rates; Average Savings Deposit Rate; 3-Month Time Deposit Rate; Average Commercial Bank Lending Rate
Money Supply	Broad Money (M2); Narrow Money (M1); Total Liquidity (M2+); Reserve Money; Currency Outside Banks; Demand Deposits; Savings & Time Deposits; Foreign Currency Deposits; Net Foreign Assets (BoG); Private Sector Credit
Economic Activity	BoG Composite Index of Economic Activity (Nominal, Real)
Commodity Prices	International Brent Crude Oil, Cocoa, and Gold prices (monthly average, US$)

Note: The full 56-variable cleaned dataset, including banking-soundness and external-sector indicators with shorter reporting histories, is documented separately in the companion data-preparation workbook (Ghana_Macro_Monetary_Data_Cleaned.xlsx) and was not used in the present modeling exercise because its shorter common window (2019–2023, 52 months) would have left too few Train-period observations for reliable hyperparameter tuning of the machine-learning models.

3.2 Feature Engineering and Leakage Control

The forecasting target is Headline Inflation, year-on-year percent, as published by the Bank of Ghana. Every predictor in the feature matrix is constructed so that the row dated t contains only information that was genuinely available at $t-1$: lagged levels and log-levels of monetary aggregates and commodity prices (lags of 1, 2, 3, 6, and 12 months), lagged policy and market interest rates, rolling 3- and 6-month means of the target computed strictly over months $t-3$ through $t-1$, a lagged month-on-month change in the target, lagged year-on-year M2 growth, and the calendar month of the forecast date (known in advance and included to capture residual seasonality, e.g. food-harvest cycles). No contemporaneous regressor, rolling window, or difference that would incorporate information from period t itself is used to predict the target at t. This is a deliberate and, in our reading of the applied machine-learning-for-macro literature, under-emphasized methodological choice: a naive contemporaneous-regressor design can produce deceptively strong in-sample fit statistics that do not correspond to any forecast a real user could have produced in real time.

After lagging (up to 12 months) and dropping rows with insufficient history, 266 monthly observations remain, spanning February 2001 to April 2023.

3.3. Model Development

To make the workflow concrete, the diagram below (Fig. 1) sketches the pipeline described in this section. Cleaned monthly data feed into the leakage-free feature matrix, which is split chronologically into training, validation, and test windows. Each candidate model is then tuned by walk-forward cross-

validation on the training window only, checked against the validation window, refit on the combined training-and-validation data, and finally scored exactly once on the untouched test window: the same sequence applied consistently to every model compared in this study.

3.3.1 Train, Validation, and Test Split

The 266-month sample is split chronologically, without shuffling, into three non-overlapping windows:

- Train (200 months, February 2001–September 2017): used to fit each model and to conduct walk-forward hyperparameter search.
- Validation (24 months, October 2017–September 2019): held out entirely from hyperparameter search; used only to compare tuned candidates on genuinely unseen data.
- Test (42 months, October 2019–April 2023): touched exactly once, after model selection was frozen. This window was deliberately chosen because it spans Ghana's 2022–23 inflation surge, during which headline inflation rose from approximately 10 percent to a peak above 50 percent, which is a genuinely difficult, policy-relevant stress test rather than a calm, favorable holdout period.

Hyperparameters for every machine-learning model are tuned via five-fold walk-forward (expanding-window) cross-validation on the Train period only, using scikit-learn's TimeSeriesSplit, with root-mean-squared error as the selection criterion. Once hyperparameters are frozen, each model is refit on Train and Validation combined and used to generate Test-period forecasts, mirroring how a forecaster would, in practice, use all data available up to the forecast date. The Test set is evaluated exactly once, after this process is complete, to avoid any form of test-set leakage into model or hyperparameter selection.

3.3.2 Candidate Models

Nine models are compared, spanning three methodological families:

- Naive baselines (Hyndman & Athanasopoulos, 2021): a Random Walk (the forecast for month t equals the observed value at t − 1) and a Seasonal Naive model (the forecast equals the observed value 12 months earlier).
- Classical econometric benchmark: an ARIMA model (Box & Jenkins, 1970) with order selected by Akaike Information Criterion via stepwise search on the Train period, re-estimated one step ahead through the Validation and Test periods as new observations become available.
- Machine learning: Ridge (Hoerl & Kennard, 1970) and Lasso (Tibshirani, 1996) regression (regularized linear models), Random Forest (Breiman, 2001) and XGBoost (Chen & Guestrin, 2016) (tree ensembles), and Support Vector Regression (Drucker et al., 1997) with a radial basis function kernel.
- Deep learning representative: a shallow feed-forward neural network (a single-hidden-layer multilayer perceptron; Rumelhart et al., 1986). A larger, more data-hungry deep architecture such as a long short-term memory (LSTM) network (Hochreiter & Schmidhuber, 1997), of the kind used successfully for the Ghanaian cedi exchange rate by Adekoya et al. (2021), was deliberately not

used for the headline-inflation target in this study: with only 200 training months, an LSTM would almost certainly overfit and would not represent a defensible modeling choice given the sample size. This is a methodological decision, not an oversight, and we return to it explicitly in the discussion and limitations.

A note on model-update parity: ARIMA above is re-estimated one step ahead as each new observation arrives through Validation and Test, whereas the machine-learning models are tuned once on Train and then refit a single time on the combined Train-and-Validation sample before generating all Test-period forecasts, without further updating as Test unfolds. This asymmetry could in principle favor ARIMA. To check its practical importance, we refit Lasso, the strongest machine-learning model in Table 5, using the same expanding-window, one-step-ahead re-estimation procedure as ARIMA, retraining after every Test-period observation. Lasso's Test RMSE improves modestly under this rolling scheme, from 3.41 to 3.11, narrowing but not closing the gap to ARIMA's 2.41. We take this as evidence that the update-parity asymmetry accounts for only part of ARIMA's advantage; the remainder appears to reflect genuine differences in how well each model class exploits the persistence in headline inflation, consistent with the interpretation offered in Section 5.1.

Table 2 provides a quick reference for all nine models: the methodological family each belongs to and the configuration ultimately selected for it after tuning. The full hyperparameter search grids over which each model was tuned are reported in Table 3.

Table 2
Candidate Models and Selected Configuration

Model	Family	Selected configuration	Reference
Random Walk	Naive baseline	Forecast(t) = Observed(t − 1); no tunable parameters	Hyndman & Athanasopoulos (2021)
Seasonal Naive	Naive baseline	Forecast(t) = Observed(t − 12); no tunable parameters	Hyndman & Athanasopoulos (2021)
ARIMA(3,1,0)	Classical econometric	Order chosen by AIC on Train; re-estimated one step ahead through Validation and Test	Box & Jenkins (1970)
Ridge	Regularized linear (ML)	α = 100.0	Hoerl & Kennard (1970)
Lasso	Regularized linear (ML)	α = 0.5	Tibshirani (1996)
Random Forest	Tree ensemble (ML)	400 trees; max depth 5; min. samples per leaf 1	Breiman (2001)
XGBoost	Gradient-boosted trees (ML)	200 trees; max depth 4; learning rate 0.1; subsample 0.8	Chen & Guestrin (2016)
Support Vector Regression	Kernel method (ML)	RBF kernel; C = 1; ε = 0.1; γ = scale	Drucker et al. (1996)
MLP (neural network)	Shallow feed-forward deep learning	1 hidden layer of 32 units; α = 0.001; learning rate = 0.001	Rumelhart et al. (1986)

Table 3
Hyperparameters Tuned (TimeSeriesSplit, 5-fold, Train period) and Final Selected Values

Model	Search space	Selected value(s)
ARIMA	p, q ∈ {0,…,5}; d selected automatically; AIC-minimizing stepwise search	order = (3, 1, 0)
Ridge	α ∈ {0.01, 0.1, 1, 10, 50, 100}	α = 100.0
Lasso	α ∈ {0.001, 0.01, 0.05, 0.1, 0.5, 1.0}	α = 0.5
Random Forest	n_estimators ∈ {200,400}; max_depth ∈ {3,5,8,None}; min_samples_leaf ∈ {1,3,5}	n_estimators = 400; max_depth = 5; min_samples_leaf = 1
XGBoost	n_estimators ∈ {200,400}; max_depth ∈ {2,3,4}; learning_rate ∈ {0.01,0.05,0.1}; subsample ∈ {0.8,1.0}	n_estimators = 200; max_depth = 4; learning_rate = 0.1; subsample = 0.8
Support Vector Regression	C ∈ {1,10,50,100}; ε ∈ {0.1,0.5,1.0}; γ ∈ {scale, auto}	C = 1; ε = 0.1; γ = scale
MLP (neural network)	hidden layer sizes ∈ {(16,), (32,), (16,8)}; α ∈ {0.001,0.01,0.1}; learning rate ∈ {0.001,0.01}	hidden layer=(32,); α = 0.001; learning rate = 0.001

3.3.3 Evaluation Metrics

Model performance is assessed on each split using four complementary metrics: root-mean-squared error (RMSE) and mean absolute error (MAE), both in percentage points of headline inflation; mean absolute percentage error (MAPE); and the coefficient of determination (R^2). RMSE is treated as the primary criterion for model ranking because it penalizes large forecast errors more heavily than MAE: an appropriate emphasis for a policy application in which large misses (such as failing to anticipate the scale of the 2022 surge) carry disproportionate real-world cost.

4. Results

4.1 Descriptive Statistics

Table 4 reports descriptive statistics for the core variables underlying the analysis over the full 280-month sample (January 2000–April 2023). Headline inflation averaged 16.45 percent with a standard deviation of 9.07 points, ranging from a low of 7.50 percent to a peak of 54.10 percent, the latter reached during the 2022–23 surge captured in the test period. Policy and market interest rates moved broadly in tandem with inflation, as expected under an inflation-targeting regime, while monetary aggregates and private-sector credit exhibit the strong upward trend and large variance characteristic of nominal series in a growing, moderately-to-highly inflationary economy: the motivation for using log-levels of these series in the feature matrix (Section 3.2).

Table 4
Descriptive Statistics of Core Variables, January 2000–April 2023 (n = 280 months)

Variable	Mean	Std. Dev.	Min	Max
Headline Inflation, YoY (%)	16.45	9.07	7.50	54.10
Monetary Policy Rate (%)	18.92	4.97	12.50	29.50
91-Day Treasury Bill Discount Rate (%)	18.95	7.82	9.04	41.85
Broad Money M2 (GHS millions)	28,719.94	35,444.51	392.86	150,496.16
Private Sector Credit (GHS millions)	16,843.11	18,036.51	379.09	73,744.05
Brent Crude Oil (US$/barrel)	65.96	29.29	19.17	134.79
Cocoa (US$/tonne)	2,213.40	659.01	701.43	3,430.35
Gold (US$/fine ounce)	1,058.68	528.78	257.82	2,000.69

4.2 Model Performance Comparison

Table 5 ranks all nine models by Test-set RMSE, the primary out-of-sample evaluation criterion. The ARIMA(3,1,0) benchmark, selected by AIC on the Train period, achieved the lowest Test RMSE (2.41 percentage points), narrowly ahead of the naive Random Walk (2.65) and Lasso regression (3.41). Every model with greater functional flexibility (Ridge, Random Forest, XGBoost, Support Vector Regression, and the neural network) recorded a higher Test RMSE than these three simplest benchmarks, and the Seasonal Naive baseline performed worst of all (RMSE 15.08, negative R^2), reflecting the fact that headline inflation in this period did not simply repeat its value from twelve months earlier.

Table 5
Test-Set Model Ranking by RMSE (42 months, October 2019–April 2023)

Rank	Model	RMSE	MAE	MAPE (%)	R^2
1	ARIMA(3,1,0)	2.41	1.34	6.75	0.971
2	Random Walk	2.65	1.66	7.75	0.965
3	Lasso	3.41	2.12	9.88	0.941
4	MLP (neural network)	5.35	3.85	22.98	0.855
5	Ridge	7.01	4.64	19.20	0.752
6	Random Forest	7.58	4.17	14.52	0.710
7	XGBoost	9.04	4.97	16.72	0.587
8	Support Vector Regression	13.37	7.94	30.57	0.097
9	Seasonal Naive	15.08	9.32	35.48	−0.148

These rankings reflect point-estimate differences in Test RMSE; to assess whether they are statistically distinguishable we conducted pairwise Diebold–Mariano tests comparing ARIMA's squared forecast errors against every other model's over the 42-month Test window. ARIMA's advantage over the Random Walk (DM = − 0.98, p = .335) and over Lasso (DM = − 1.53, p = .135) is not statistically significant at conventional levels, so the fine ranking among these three closest competitors in Table 5 should be read as suggestive rather than conclusive. By contrast, ARIMA significantly outperforms Ridge (DM = − 3.49, p = .001), Random Forest (DM = − 2.55, p = .015), XGBoost (DM = − 2.73, p = .009), Support Vector Regression (DM = − 3.16, p = .003), the MLP (DM = − 2.86, p = .007), and the Seasonal Naive baseline (DM = − 3.43, p = .001). Put differently, ARIMA and the Random Walk are statistically indistinguishable from one another on this Test window, and both are reliably better than the more flexible machine-learning models; it is that latter contrast, not the fine distinction between ARIMA and the Random Walk, that is the central and statistically robust finding of this paper.

Figure 2 lines up the actual inflation path against every model's Test-period forecast, and the picture matches the ranking in Table 5 closely. The dashed ARIMA and Random Walk lines track the solid black actual line most closely, following the climb through 2022 with only a short lag, while the machine-learning models drift further away as inflation accelerates; most land below the true peak and take longer to catch back up once inflation starts falling. Every model tracks the series reasonably well in the calmer years before 2022; the gap only really opens up once the surge begins, which is exactly the period this study set out to stress-test. At the single hardest point in the window (the December 2022 peak, where actual inflation reached 54.1 percent), ARIMA's forecast (54.3) and the Random Walk's (50.3) remain the closest of any model; the nearest machine-learning competitor, Lasso, is over 8 points off (45.7), so no machine-learning model overtakes the classical benchmarks even momentarily at the peak.

Figure 3 puts the same comparison in bar-chart form across all three error metrics. ARIMA and the Random Walk take the top two spots on every metric, and the picture holds together throughout: RMSE, MAE, and MAPE all tell the same story, with the Seasonal Naive baseline and Support Vector Regression trailing furthest behind.

4.3 Overfitting Diagnostics: Train versus Test Performance

Table 6 compares each model’s RMSE on the Train, Validation, and Test splits. The gap between Train and Test performance is the clearest diagnostic of overfitting in this exercise. XGBoost fits the Train period almost perfectly (RMSE 0.03) yet degrades sharply out of sample (Test RMSE 9.04): a difference of two orders of magnitude that is the textbook signature of a flexible model memorizing training-period idiosyncrasies rather than learning generalizable structure. Random Forest shows a similarly large, if less extreme, gap (0.83 to 7.58). By contrast, ARIMA’s Train and Test RMSE (3.23 and 2.41 respectively) are of comparable magnitude, and the Random Walk baseline is, by construction, not tunable and therefore cannot overfit in this sense. The MLP neural network is a partial exception: its Validation RMSE (7.61) is markedly worse than its Test RMSE (5.35), suggesting the Validation window, which includes the onset of COVID-19-related macro volatility in 2020, was itself an unusually difficult sub-period for this

specification, even though its Test-period generalization remains substantially worse than the classical benchmarks.

The near-zero XGBoost training RMSE is striking enough to warrant a direct check for target leakage rather than a benign explanation. We re-examined the feature-construction pipeline (Section 3.2) for every predictor and confirmed that each one is either a genuine lag of at least one month, a rolling statistic computed strictly on already-lagged values, or the calendar month, which is known in advance; no predictor is drawn from the same period as the target. We also computed the correlation between every engineered feature and the contemporaneous target: the strongest is the target's own first lag, at 0.978, which is consistent with headline inflation's high persistence and not indicative of leakage. We therefore attribute XGBoost's near-perfect training fit to model capacity: 200 boosting rounds of depth-4 trees applied to only 200 training months gives the model ample room to fit training-period idiosyncrasies, rather than to a data-construction error.

Table 6
RMSE by Model and Split (percentage points of Headline Inflation)

Model	Train RMSE	Validation RMSE	Test RMSE	Test − Train Gap
Random Walk	1.62	0.53	2.65	+ 1.03
Seasonal Naive	9.12	2.40	15.08	+ 5.96
ARIMA(3,1,0)	3.23	0.51	2.41	−0.82
Ridge	1.67	0.59	7.01	+ 5.34
Lasso	1.63	0.96	3.41	+ 1.78
Random Forest	0.83	0.79	7.58	+ 6.75
XGBoost	0.03	0.73	9.04	+ 9.01
Support Vector Regression	3.12	4.84	13.37	+ 10.25
MLP (neural network)	1.80	7.61	5.35	+ 3.55

4.4 Feature Importance

Figure 4 reports the top twelve feature importances for the two tree-ensemble models, Random Forest and XGBoost. In both models, the single most important predictor by a wide margin is the target's own first lag (Headline Inflation at t − 1), consistent with the high persistence documented in Section 4.1 and with the strong performance of the Random Walk baseline. Beyond the first lag, both models assign meaningful importance to the second lag of inflation, the 3-month rolling mean of inflation, and lagged Treasury bill discount rates: variables that proxy for the current monetary policy stance and the recent inflation trend. Lagged international Brent crude oil prices (at a 12-month lag) also register as informative in both models, consistent with the pass-through of global energy prices into Ghanaian consumer prices documented in the literature (Kunawotor et al., 2021).

4.5 Residual Diagnostics for the Best-Performing Model

Figure 5 looks at how ARIMA's errors behave on the Test set, since it was the best-performing model overall. For most of the period the errors stay small and hover close to zero, which is what you would hope to see. The biggest misses show up around the steepest months of the 2022 inflation spike, which is not surprising, since a one-step-ahead model can only really catch up to a sharp turn after it has already started. The error distribution is roughly bell-shaped and centered near zero, without any obvious skew, and plotting the errors against the forecasts does not reveal any strong pattern either, though errors do tend to get a little larger when inflation itself is higher. That is a common feature of inflation data during high-inflation episodes, and it is worth keeping in mind: a follow-up model aimed at producing forecast ranges, not just single-number forecasts, would probably want to account for that changing variance directly (for example with a GARCH-type specification) rather than assuming the same margin of error applies at 10 percent inflation and at 50 percent.

5. Discussion

5.1. Key Findings

This study set out to answer a simple question: for Ghana's headline inflation, do machine learning and deep learning models actually forecast better than the classical tools already in wide use, once they are tested fairly and under real pressure? The results below speak to that question directly.

The central empirical finding of this study (that a classical ARIMA benchmark and a naive random walk outperform every machine-learning and deep-learning model tested, out of sample, through Ghana’s 2022–23 inflation surge) runs counter to a common presumption in the applied machine-learning-for-macroeconomics literature, including studies of closely related Ghanaian and Nigerian series that report favorable results for flexible models (Adekoya et al., 2021; Akande et al., 2021). We view this not as a contradiction of that literature but as a complementary and clarifying result, for three reasons.

First, the target series matters. Adekoya et al. (2021) forecast an exchange rate, a series with different persistence and volatility dynamics than headline year-on-year inflation, which by construction embeds a twelve-month moving comparison and is consequently far more autocorrelated. A highly persistent series rewards models, like ARIMA and the random walk, that are explicitly built to exploit that persistence; it does not automatically reward models with the greatest functional flexibility. Second, sample size matters. With 200 training months, tree ensembles, kernel methods, and neural networks have ample capacity to fit training-period noise that does not generalize: precisely the pattern documented in Table 6, where XGBoost’s Train RMSE of 0.03 collapses to a Test RMSE of 9.04. Akande et al. (2021) evaluate their stacked ensemble primarily on in-sample decomposition of inflation drivers rather than a genuinely held-out, crisis-spanning forecast window; our results suggest that the same class of models, when subjected to strict one-step-ahead, out-of-sample evaluation through a structural break, does not retain its apparent advantage. Third, and most importantly for policy, the test period itself

contains a regime shift that none of the models had seen the likes of in training. A structural break is, definitionally, the hardest case for any statistical model, classical or machine-learning, to forecast, but our results suggest that simpler, more constrained models degrade more gracefully under this kind of stress than more flexible ones, which appear to overfit spurious relationships from the pre-crisis training period that break down precisely when they matter most.

These findings carry a direct implication for forecasting practice at the Bank of Ghana and comparable institutions: model selection should not default to the most sophisticated available method. A well-specified, regularly re-estimated ARIMA model, cheap to maintain, transparent to interpret, and robust to the sample-size constraints inherent in monthly macro data, remains a strong and, in this evaluation, superior choice for one-step-ahead headline inflation forecasting relative to the machine-learning alternatives tested here. This is consistent with the broader inflation-targeting literature's emphasis on transparency and communicability of the forecasting process to policy committees and the public (Svensson, 2010; Chugunov et al., 2019).

This does not imply that machine learning has no role in Ghanaian inflation forecasting. Several extensions plausibly favor ML approaches: richer, higher-frequency, or textual/news-based feature sets (of the kind envisioned by Tassi, 2019, in the broader big-data-for-economics literature) may contain nonlinear signal that a linear or low-order ARIMA specification cannot capture; regime-aware architectures (Markov-switching models, or ensemble methods that explicitly detect and adapt to structural breaks) may generalize better than the static specifications used here; and hybrid approaches that use a classical model for the persistent trend and a machine-learning model for the residual, higher-frequency component may combine the robustness of the former with the flexibility of the latter. We view these as promising directions rather than as claims already substantiated by the present results.

Two pieces of related work help place these findings in context. First, our results are consistent with the small-sample advantage for classical methods documented by Cerqueira et al. (2019): with roughly 200 training months, flexible machine-learning models are prone to overfitting and can lag well-specified classical benchmarks, a pattern their cross-domain analysis attributes to learning curves that only favor machine learning once the effective sample size is considerably larger than what is available for a single country's monthly inflation history. A learning-curve analysis, testing whether machine learning begins to close the gap as more data, higher-frequency covariates, or richer exogenous inputs become available, would directly test this explanation for Ghana and is a natural next step. Second, a recent study applying machine learning to Ghanaian inflation directly, Baidoo and Obeng (2024), reports favorable results for machine-learning models. The apparent tension with our findings is, we think, informative rather than contradictory: that study does not adopt as strict a leakage-free, one-step-ahead design, does not hold out a genuinely adversarial crisis period for final evaluation, and does not benchmark against baselines as demanding as the random walk used here. Read together, the two studies suggest that machine learning's apparent advantage for Ghanaian inflation is sensitive to evaluation design: precisely the methodological point this paper set out to make.

5.2. Practical and Policy Implications

For a forecasting team at the Bank of Ghana, the practical takeaway is not to chase the newest modeling technique for its own sake. ARIMA is cheap to re-estimate every month, easy to explain to a monetary policy committee that needs to understand why a forecast moved, and it held up better than every machine-learning alternative tested here, including through the hardest stretch of the sample. That combination (low cost, transparency, and demonstrated robustness) is hard to beat for an institution that has to publish and defend its forecasts in real time.

That said, machine learning still has a place in the toolkit, just not as a wholesale replacement. A sensible approach would be to keep ARIMA, or a similar classical model, as the primary forecast, and use machine-learning models as a cross-check or an ensemble input rather than the headline number, flagging cases where a data-driven model disagrees sharply with the classical forecast as worth a closer look, rather than trusting the more complex model by default. Before any machine-learning model goes into operational use for policy forecasting, it should be tested the way this study tested one: on a genuinely held-out period that includes a stress episode, not just a calm stretch of history.

5.3. Limitations and Future Research

Several limitations qualify these conclusions and point toward future research. First, the training sample, while long by the standards of Ghanaian monetary statistics (200 months), remains modest relative to the data requirements of the most data-hungry deep-learning architectures; a larger deep-learning model such as an LSTM was deliberately excluded from this comparison for that reason, and our conclusions about deep learning should be read as applying to the shallow feed-forward specification tested, not as a categorical claim about all deep architectures. Second, the study is confined to a single country and a single target series (headline year-on-year inflation); the relative performance of classical and machine-learning models may differ for food or core inflation components, for shorter forecast horizons, or for other Sub-Saharan African economies with different inflation dynamics. Third, the feature set, while broad (24 core macro-monetary variables with extensive lag structures), is limited to Bank of Ghana's own published statistics and international commodity prices; it does not incorporate survey-based inflation expectations, fiscal high-frequency indicators, or textual/news-based signals, any of which might alter the relative standing of flexible machine-learning models. Fourth, our evaluation focuses on point-forecast accuracy (RMSE, MAE, MAPE, R^2); a fuller policy-relevant evaluation would also assess the calibration of forecast intervals, particularly given the heteroskedasticity suggested by the residual diagnostics in Section 4.5, and would account for the asymmetric costs of over- versus under-predicting inflation emphasized by Charemza et al. (2014).

A fifth limitation concerns the breadth of the classical benchmark set and the robustness of a single evaluation split. We compare machine-learning models against one classical anchor, ARIMA, and do not evaluate dynamic regressions with exogenous variables (ARIMAX), unobserved-components or time-varying-parameter models, Markov-switching autoregressions, or Bayesian shrinkage regressions; adaptive shrinkage priors such as the Horseshoe have shown strong short-horizon inflation-forecasting performance in other emerging-market settings when paired with a rich predictor set (Taveeapiradeecharoen & Arwatchanakarn, 2025), and represent a natural next comparator beyond the

Ridge and Lasso specifications used here. Likewise, all results in this paper come from a single chronological Train/Validation/Test split; a full rolling-origin backtest across multiple test windows, paired with Diebold–Mariano tests at each origin, would give a more complete picture of how sensitive the ranking in Table 5 is to the particular split used. Finally, as noted above, we do not evaluate density forecasts or interval calibration (e.g., CRPS, pinball loss, or coverage of prediction intervals), which would matter directly for a policy committee weighing forecast uncertainty rather than point forecasts alone.

Future research should pursue three directions suggested by these results. First, regime-aware or structural-break-adaptive models, such as Markov-switching specifications, or ensemble methods with explicit change-point detection, should be evaluated against the same leakage-free, crisis-spanning test design used here, to test directly whether regime-awareness rather than raw functional flexibility is what machine learning needs to add value for this series. Second, hybrid architectures that combine a classical model for the persistent, trend component of inflation with a machine-learning model for the residual, higher-frequency component merit direct testing. Third, extending the same forecasting design to food and core inflation sub-components, and to a panel of West African economies, would clarify whether the advantage of classical models documented here for Ghana's aggregate headline inflation generalizes regionally or is a feature of this particular series and episode.

6. Conclusions

This study rigorously evaluated whether modern machine learning and deep learning architectures offer a genuine, policy-relevant forecasting advantage over classical econometric benchmarks for Ghana's headline year-on-year inflation within an operational central banking context. Utilizing 280 months (2000–2023) of Bank of Ghana macro-monetary data, we constructed a strictly leakage-free, one-step-ahead forecasting matrix and evaluated nine competing specifications across a chronologically held-out 42-month test window (October 2019–April 2023) deliberately selected to span Ghana's severe 2022–2023 inflation surge above 50 percent. Contrary to the presumption that greater functional flexibility yields superior predictive accuracy, the classical ARIMA(3,1,0) benchmark achieved the lowest test-set root-mean-squared error (2.41 percentage points), narrowly outperforming the naive Random Walk baseline (2.65) and regularized Lasso regression (3.41), while every complex nonlinear machine learning specification—including Random Forest, XGBoost, Support Vector Regression, and a feed-forward neural network—recorded substantially higher forecast errors. Diebold–Mariano tests confirmed that while ARIMA, the Random Walk, and Lasso are statistically indistinguishable from one another on this test set, ARIMA's performance advantage over all nonlinear machine learning models is statistically significant. Overfitting diagnostics revealed that flexible tree ensembles fit training-period noise near-perfectly but degraded sharply when confronting the test period's structural break, illustrating that high-capacity algorithms are particularly prone to fitting spurious relationships in modest sample sizes (200 training months) that fail to generalize during macroeconomic crises. Because headline year-on-year inflation is inherently characterized by strong autoregressive persistence, simple specifications designed to exploit this persistence degrade far more gracefully under extreme stress than unconstrained non-linear models. For monetary policy authorities such as the Bank of Ghana, these findings demonstrate that

operational forecasting should not default to algorithmic complexity; rather, well-specified, low-cost, and transparent classical models remain the most reliable primary forecasting anchors, with machine learning best positioned as an auxiliary cross-check rather than a primary tool. Naturally, these conclusions come with real limitations: the monthly sample size remains modest for data-hungry deep learning architectures (such as LSTMs), the feature set relied exclusively on published macro-monetary statistics without survey-based expectations or high-frequency signals, the evaluation focused on point-forecast metrics rather than density or interval calibration, and the analysis was confined to a single country's aggregate headline inflation. To extend this line of inquiry, future research should investigate regime-aware models equipped with explicit change-point detection, hybrid architectures combining classical trend components with algorithmic residuals, disaggregated food and core inflation sub-components, and multi-country regional panels across Sub-Saharan Africa to establish whether these findings generalize across broader emerging market settings.

Declarations

CRediT authorship contribution statement

Conceptualization, I.N., B.O., A.B. and G.O.F.; methodology, I.N., B.O., A.B. and G.O.F.; software, G.O.F.; validation, I.N., B.O., A.B. and G.O.F.; formal analysis, I.N., B.O., A.B. and G.O.F.; investigation, I.N., B.O., A.B. and G.O.F.; resources, I.N., B.O., A.B. and G.O.F.; writing—original draft preparation, I.N., B.O., A.B. and G.O.F; writing—review and editing, I.N., B.O., A.B. and G.O.F.; supervision, G.O.F. All authors have read and agreed to the publication of the manuscript.

Conflicts of Interest:

The authors declare no conflicts of interest.

Ethics approval statement:

Not applicable

Funding:

This research received no external funding.

Acknowledgement

The authors would like to thank the Kwame Nkrumah University of Science and Technology, Kumasi (KNUST), University of Skills Training and Entrepreneurial Development (USTED), Kumasi, Faculty of Health Studies, University of Bradford, and the Department of Forest Engineering, Forest Management

Planning, and Terrestrial Measurements, Faculty of Silviculture and Forest Engineering, Transilvania University of Brasov, for providing some of the resources needed for this study.

Data Availability

The dataset and Python script used for data analysis and visualisation are available as supplementary files on the figshare repository at: https://figshare.com/s/992511ca2a97e76f7d6d.

References

1. Acquah-Sam, E. (2017). Influencers of Inflation in Ghana. *European Scientific Journal ESJ*, *13*(7), 140. https://doi.org/10.19044/esj.2017.v13n7p140
2. Adekoya, A. F., Nti, I. K., & Weyori, B. A. (2021). Long short-term memory network for predicting exchange rate of the Ghanaian cedi. *FinTech*, *1*(1), 25–43. https://doi.org/10.3390/fintech1010002
3. Akande, E., Akanni, E., Taiwo, O. F., Joshua, J. D., & Anthony, A. (2021). *Predicting inflation component drivers in Nigeria: A stacked ensemble approach* [Preprint]. https://doi.org/10.21203/rs.3.rs-1080209/v1
4. Andrle, M., Berg, A., Morales, R. A., Portillo, R., & Vlcek, J. (2013). Forecasting and monetary policy analysis in low-income countries; food and non-food inflation in Kenya. *IMF Working Papers*, *13*(61), 1. https://doi.org/10.5089/9781475537444.001
5. Anwar, S., & Islam, I. (2013). Should developing countries target low, single-digit inflation to promote growth and employment? In *Beyond macroeconomic stability* (pp. 98–130). Palgrave Macmillan UK. https://doi.org/10.1057/9781137379252_4
6. Ayalew, Z. (2022). Dynamic interaction between saving, investment and economic growth in Ethiopia. *Journal of Economics and Sustainable Development*. https://doi.org/10.7176/jesd/13-9-04
7. Badia, M. M. (2016). Fiscal Monitor, October 2016: Debt, use it wisely. *International Monetary Fund*. https://doi.org/10.5089/9781513560601.089
8. Baidoo, T. G., & Obeng, A. (2024). Navigating Inflation in Ghana: How Can Machine Learning Enhance Economic Stability and Growth Strategies? arXiv preprint arXiv:2410.05630. https://doi.org/10.48550/arXiv.2410.05630
9. Box, G. E. P., & Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day.
10. Breiman, L. (2001). Random Forests. Machine Learning, 45. https://doi.org/10.1023/A:1010933404324
11. Cerqueira, V., Torgo, L., & Soares, C. (2019). Machine Learning vs Statistical Methods for Time Series Forecasting: Size Matters. arXiv preprint arXiv:1909.13316. https://doi.org/10.48550/arXiv.1909.13316
12. Charemza, W., Kharin, Y., & Maevskiy, V. (2014). Bilinear forecast risk assessment for non-systematic inflation: Theory and evidence. In *Dynamic modeling and econometrics in economics and finance* (pp.

205–232). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-42039-9_6

13. Chen, T., Guestrin, C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and, & Mining, D. (2016). https://doi.org/10.1145/2939672.2939785
14. Chugunov, I., Pasichnyi, M., & Nepytaliuk, A. (2019). Macroeconomic effects of inflation targeting in advanced and emerging market economies. *Banks and Bank Systems*, *14*(4), 153–165. https://doi.org/10.21511/bbs.14(4).2019.15
15. Chugunov, I., Pasichnyi, M., Kaneva, T., Nepytaliuk, A., & Koroviy, V. (2021). The influence of inflation targeting on economic growth in the OECD and developing countries. *Montenegrin Journal of Economics*, *17*(2), 173–186. https://doi.org/10.14254/1800-5845/2021.17-2.14
16. Cust, J., & Mihalyi, D. (2017). *Evidence for a presource curse? Oil discoveries, elevated expectations, and growth disappointments* (World Bank Policy Research Working Paper No. 8140). World Bank. https://doi.org/10.1596/1813-9450-8140
17. Dagume, A. M. (2022). Exchange rate volatility and macroeconomic variables in South Africa. *International Journal of Economics and Financial Issues*, *12*(6), 1–14. https://doi.org/10.32479/ijefi.13446
18. Drucker, H., Burges, C. J., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support vector regression machines. *Advances in neural information processing systems*, *9*.
19. Ejem, C. A., & Ogbonna, U. G. (2020). Financial conditions index and economic performance in Nigeria. *American Finance & Banking Review*, *5*(1), 62–70. https://doi.org/10.46281/amfbr.v5i1.564
20. Filho, F. F., & Milan, M. (2018). Excess real interest rates and the inflation targeting regime in Brazil: Monetary policy ineffectiveness and rentiers' interests. *Applied Economics and Finance*, *5*(6), 84–100. https://doi.org/10.11114/aef.v5i6.3710
21. Forbes, K. J., Hjortsoe, I., & Nenova, T. (2017). Shocks versus structure: Explaining differences in exchange rate pass-through across countries and time. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2999637
22. Frankel, J. A. (2017). How to cope with volatile commodity export prices: Four proposals. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3025565
23. George-Anokwuru, C. C. (2023). Monetary policy and misery index in Nigeria. *European Journal of Economic and Financial Research*, *7*(2). https://doi.org/10.46827/ejefr.v7i2.1494
24. Godfrey, O. U., & Agwu, E. C. (2020). Economic performance and financial conditions index nexus: Evidence from selected Sub-Saharan African countries. *Journal of Finance and Accounting*, *8*(3), 158. https://doi.org/10.11648/j.jfa.20200803.17
25. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
26. Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics. https://doi.org/10.1080/00401706.1970.10488634

27. Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts.
28. International Monetary Fund. (2013). Uganda: First review under the policy support instrument. *(IMF Staff Country Reports No*. https://doi.org/10.5089/9781475522754.002. 13/375).
29. International Monetary Fund. (2018). Uganda: Technical assistance report—Monetary and foreign exchange operations, recapitalization, and act revision. *(IMF Staff Country Reports No*. https://doi.org/10.5089/9781484337813.002. 18/10).
30. International Monetary Fund, Fiscal Affairs Department. (2018). Fiscal Monitor, April 2018. *International Monetary Fund*. https://doi.org/10.5089/9781484333952.089
31. International Monetary Fund, Monetary and Capital Markets Department. (2020a). Global Financial Stability Report, April 2020. *International Monetary Fund*. https://doi.org/10.5089/9781513529196.082
32. International Monetary Fund, Monetary and Capital Markets Department. (2020b). Global Financial Stability Report, October 2020. *International Monetary Fund*. https://doi.org/10.5089/9781513554228.082
33. Islam, T. U., & Ahmed, D. (2023). Inflation targeting: A time-frequency causal investigation. *PLOS ONE*, *18*(12), e0295453. https://doi.org/10.1371/journal.pone.0295453
34. Jackson, E. A., Tamuke, E., Jabbie, M., & Ngombu, A. (2020). Adoption of inflation targeting in Sierra Leone: An empirical discourse. *Journal of Economic Policy Researches*, *7*(2), 21–50. https://doi.org/10.26650/jepr735604
35. Kunawotor, M. E., Bokpin, G. A., Asuming, P. O., & Amoateng, K. A. (2021). The impacts of extreme weather events on inflation and the implications for monetary policy in Africa. *Progress in Development Studies*, *22*(2), 130–148. https://doi.org/10.1177/14649934211063357
36. Madhou, A., Sewak, T., Moosa, I. A., Ramiah, V., & Gerth, F. (2021). Towards full-fledged inflation targeting monetary policy regime in Mauritius. *Journal of Risk and Financial Management*, *14*(3), 126. https://doi.org/10.3390/jrfm14030126
37. Mikailu, F., et al. (2026). Spatial dynamics, convergence, and persistence of inflation in post-rebasing Nigeria: Evidence from state-level CPI indices (2024–2025). *Hensard Journal of Policy Economics and Development*, *1*(1). https://doi.org/10.65757/hjpedv1n15
38. Muhammed-Gani, S. H., Nansir, A. M. M., & Ibrahim, A. (2022). Analysis of inflation convergence in the Economic Community of West African States (ECOWAS) zone (1987–2021). *International Journal of Research Publications*, *102*(1). https://doi.org/10.47119/ijrp1001021620223328
39. Ncanywa, T., & Ralarala, O. (2020). The effects of expectation inflation on aggregate demand: Examining the South African inflation expectation channel of South African monetary policy. *Journal of Reviews on Global Economics*, *9*, 437–445. https://doi.org/10.6000/1929-7092.2020.09.41
40. Ngarambe, E., Shukla, & Hitayezu (2016). Monetary policy and inflation control in Rwanda: 2006–2015. *International Journal of Advanced Research*, *4*(8), 1085–1096. https://doi.org/10.21474/ijar01/1308

41. Nguyen, H. (2021). *Fiscal and monetary policies in small open economies* [Doctoral dissertation, Victoria University of Wellington]. https://doi.org/10.26686/wgtn.14864796
42. Nyamao, N. R. (2014). A multiplicative autoregressive integrated moving average model for Kenya's inflation (2000:1–2013:12). *Science Journal of Applied Mathematics and Statistics*, *2*(6), 122. https://doi.org/10.11648/j.sjams.20140206.14
43. Ohemeng, W., Agyapong, E. K., & Ofori-Boateng, K. (2021). Exchange rate and inflation dynamics: Does the month or quarter of the year matter? *SN Business & Economics*, *1*(6), 1–24. https://doi.org/10.1007/s43546-021-00074-5
44. Oladipo, O. (2017). Inflation targeting and exchange rate pass-through to domestic prices: Evidence from South Africa. *Journal of Economic & Financial Studies*, *5*(5), 1–11. https://doi.org/10.18533/jefs.v5i05.294
45. Osei-Dwomoh, E., & Forkuo, G. O. (2026). Artificial intelligence in African economics: a systematic review of performance and ethical risks. *Future Business Journal*, *12*(1), 84. https://doi.org/10.1186/s43093-026-00776-y
46. Otu, O. A., George, A. O., Jude, O., Ifeyinwa, M. H., & Andrew, I. I. (2014). Application of SARIMA models in modelling and forecasting Nigeria's inflation rates. *American Journal of Applied Mathematics and Statistics*, *2*(1), 16–28. https://doi.org/10.12691/ajams-2-1-4
47. Peychinov, T., Karaivanova, A., & Mecheva, T. (2025). Predicting Traffic Load Data: ARIMA and SARIMA Comparison. *Engineering Proceedings*, *100*(1), 29. https://doi.org/10.3390/engproc2025100029
48. Pooter, M. D., Robitaille, P. T., Walker, I. A., & Zdinak, M. (2014). Are long-term inflation expectations well anchored in Brazil, Chile and Mexico? *International Finance Discussion Papers*, *2014*(1098), 1–47. https://doi.org/10.17016/ifdp.2014.1098
49. Prasad, A., Goswami, A., & Gurtu, A. (2023). Influence of demonetization on various sectors of the Indian economy. *Strategic Management*, *28*(4), 15–30. https://doi.org/10.5937/straman2200032p
50. Rossouw, J., Padayachee, V., & Bosch, A. (2011). A comparison of inflation expectations and inflation credibility in South Africa: Results from survey data. *South African Journal of Economic and Management Sciences*, *14*(3), 263–281. https://doi.org/10.4102/sajems.v14i3.8
51. Rossouw, J., Joubert, F., & Padayachee, V. (2013). An international comparison of inflation credibility surveys. *South African Journal of Economic and Management Sciences*, *16*(2), 142–153. https://doi.org/10.4102/sajems.v16i2.316
52. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning Representations by Back-Propagating Errors. *Nature*. https://doi.org/10.1038/323533a0
53. Rybinski, K., & Krynska, U. (2010). Global reserves management. In *Central bank reserves and sovereign wealth management* (pp. 3–40). Palgrave Macmillan UK. https://doi.org/10.1057/9780230250819_1
54. Svensson, L. E. O. (2010). *Inflation targeting* (NBER Working Paper No. 16654). National Bureau of Economic Research. https://doi.org/10.3386/w16654

55. Tadesse, T., & Melaku, T. (2019). Analysis of the relative impact of monetary and fiscal policies on economic growth in Ethiopia, using ARDL approach to co-integration: Which policy is more potent? *Copernican Journal of Finance & Accounting*, *8*(2), 87–115. https://doi.org/10.12775/cjfa.2019.010
56. Tassi, P. (2019). Introduction – The contributions of big data. *Économie et Statistique / Economics and Statistics*, 5–15. https://doi.org/10.24187/ecostat.2018.505d.1963
57. Taveeapiradeecharoen, P., & Arwatchanakarn, P. (2025). Forecasting Thai Inflation from Univariate Bayesian Regression Perspective. arXiv preprint arXiv:2505.05334. https://doi.org/10.48550/arXiv.2505.05334
58. Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B*. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
59. World Bank. (2012). *South Africa economic update, July 2012*. World Bank. https://doi.org/10.1596/26076
60. World Bank. (2020a). *Nigeria development update, June 2020*. World Bank. https://doi.org/10.1596/34046
61. World Bank. (2020b). *Rwanda economic update, January 2020*. World Bank. https://doi.org/10.1596/33247
62. World Bank Group. (2020). *Commodity markets outlook, April 2020*. World Bank. https://doi.org/10.1596/33624
63. Xu, S., Qamruzzaman, & Adow, A. H. E. (2021). Is financial innovation bestowed or a curse for economic sustainability: The mediating role of economic policy uncertainty. *Sustainability*, *13*(4), 2391. https://doi.org/10.3390/su13042391
64. Yabu, N., & Kessy, N. J. (2015). Appropriate threshold level of inflation for economic growth: Evidence from the three founding EAC countries. *Applied Economics and Finance*, *2*(3), 127–144. https://doi.org/10.11114/aef.v2i3.966

Figures

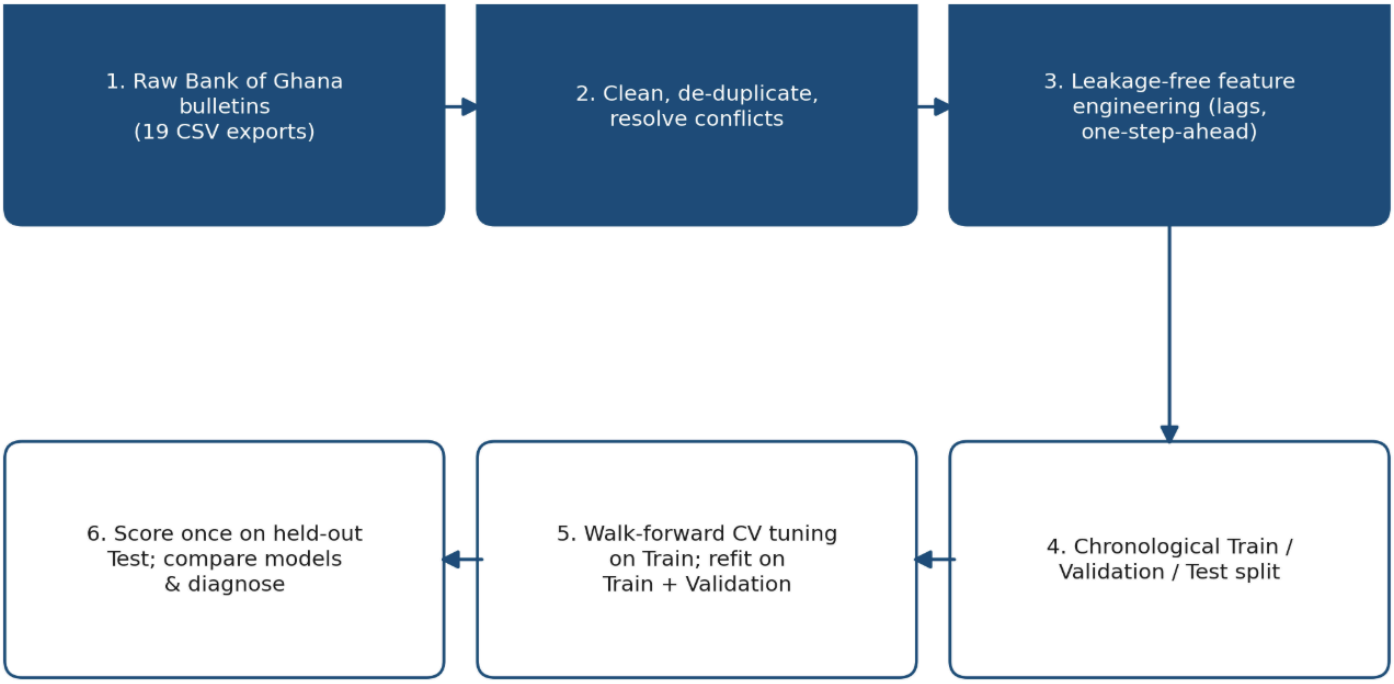


Figure 1

Pipeline overview: from raw Bank of Ghana data to the final out-of-sample evaluation.

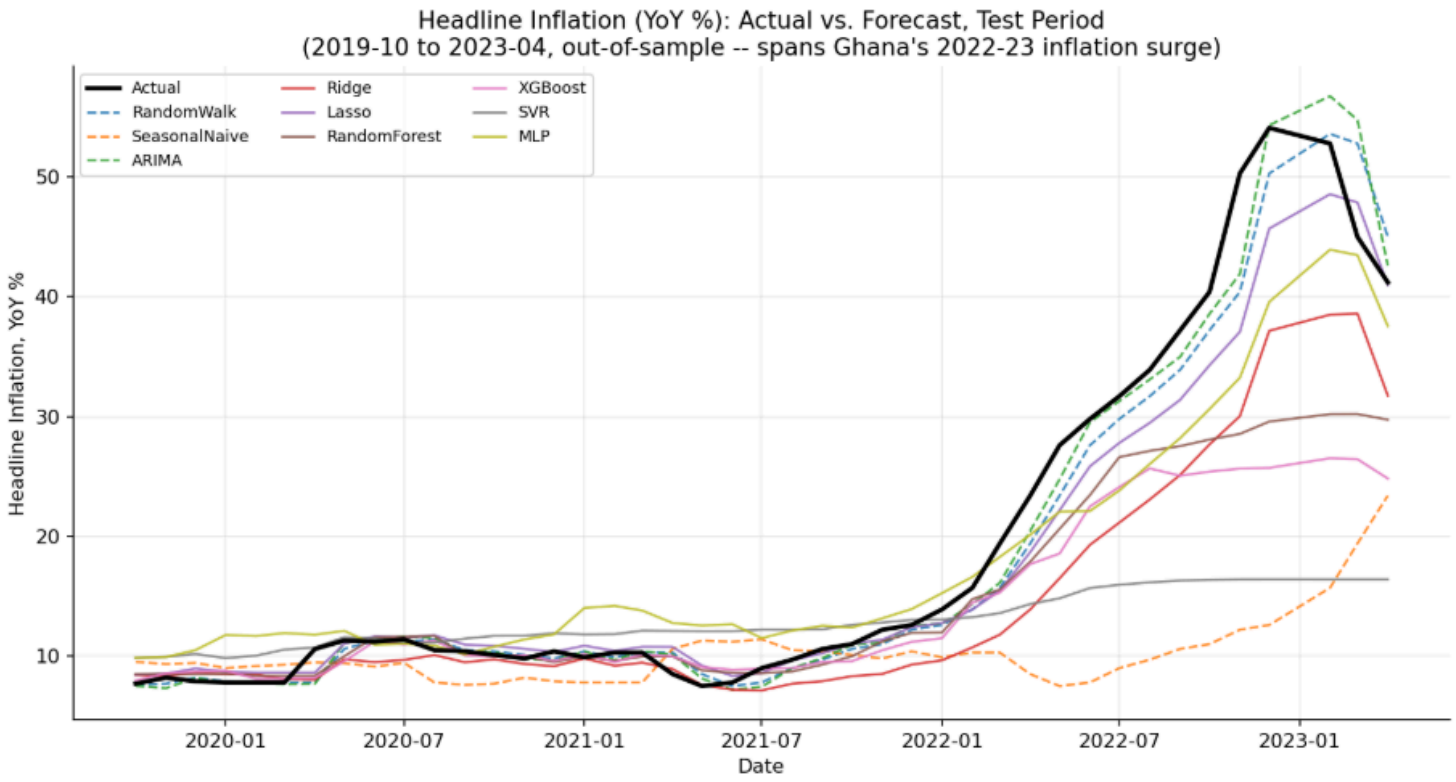


Figure 2

Actual versus forecast Headline Inflation (YoY %), Test period (Oct 2019–Apr 2023). Dashed lines denote the baseline and classical models (Random Walk, Seasonal Naive, ARIMA); solid lines denote machine-learning models.

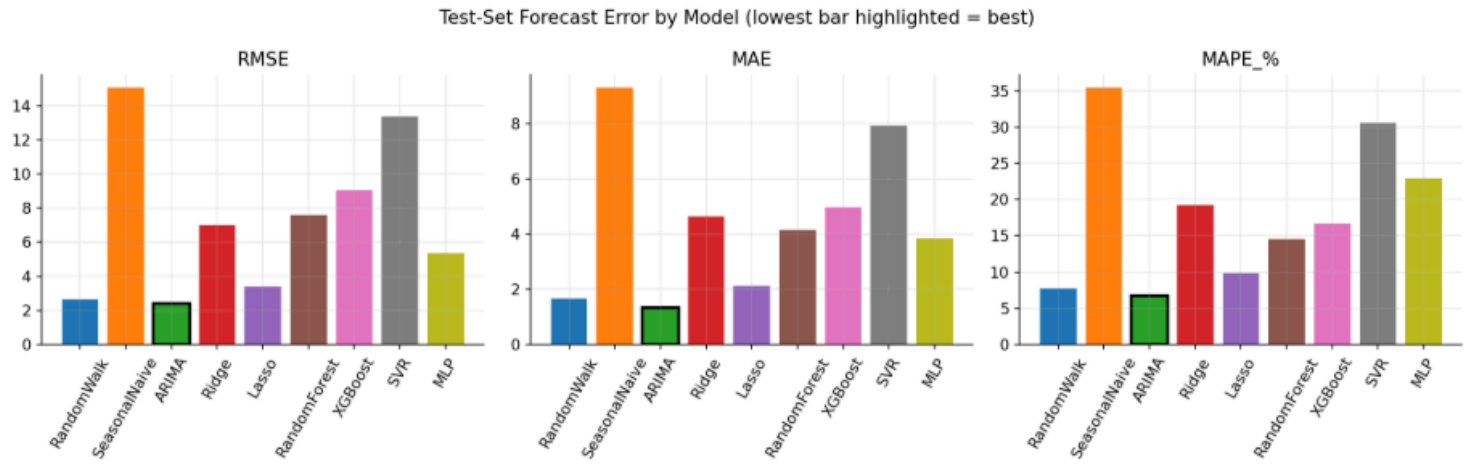


Figure 3

Test-set forecast error by model (RMSE, MAE, MAPE); the outlined bar in each panel marks the best-performing model on that metric.

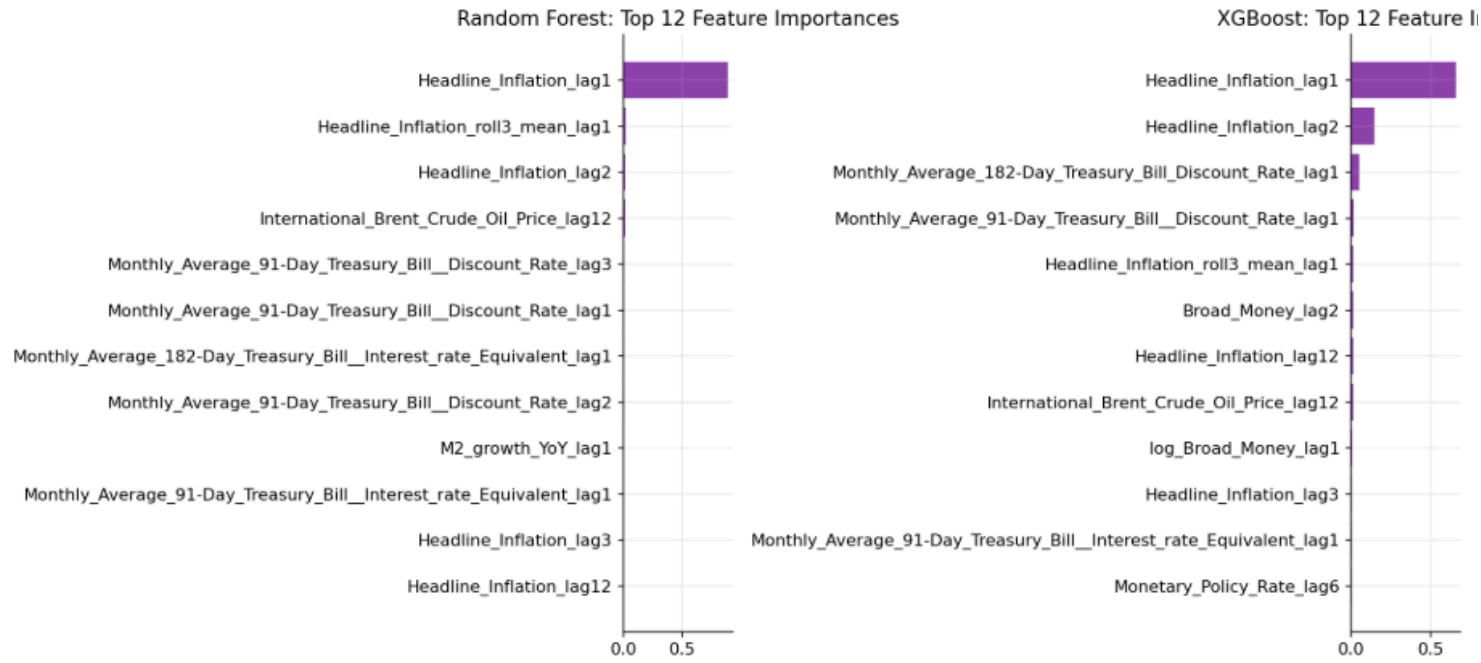


Figure 4

Top 12 feature importances, Random Forest (left) and XGBoost (right).

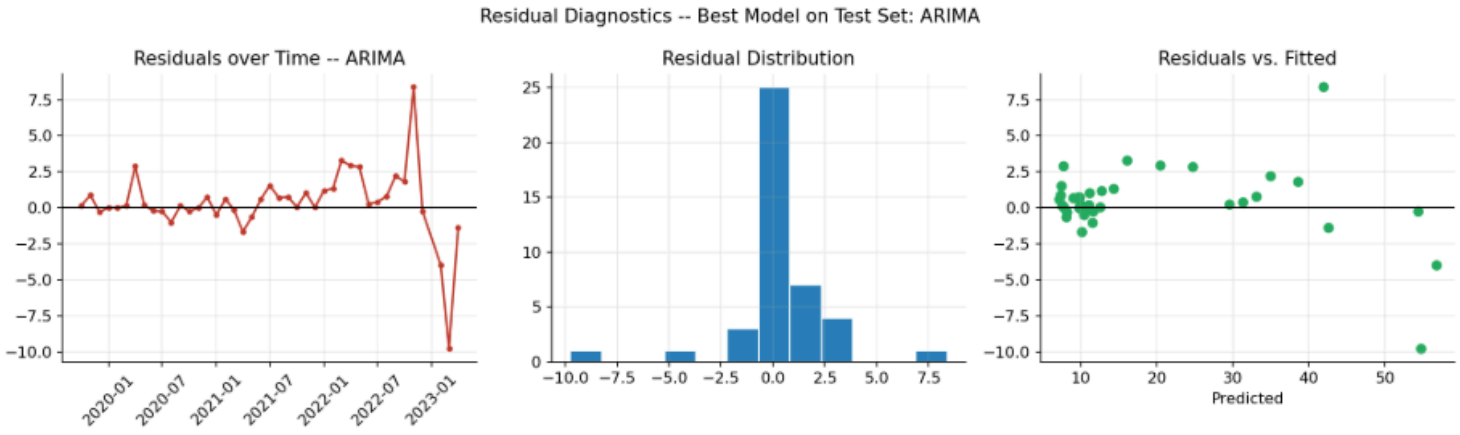


Figure 5

Residual diagnostics for the best-performing model (ARIMA) on the Test set: residuals over time (left), residual distribution (center), residuals versus fitted values (right).