CONFERENCE PROCEEDINGS/FULL PAPERS
ISBN: 978-625-93894-1-7/October 2025

"42nd RSEP International Conference on Economics, Finance and Business" 19-20 September 2025, Hotel Acta Piramides, Madrid, Spain"

How well do machine learning models forecast inflation? A study for Azerbaijan

Vugar Rahimov
PhD candidate in Econometrics, Azerbaijan State University of Economics (UNEC), Baku, Azerbaijan
vugar.m.rahimov@gmail.com

DOI: https://doi.org/10.19275/RSEPCONFERENCES362

Abstract

This study evaluates the forecasting performance of machine learning (ML) models in predicting inflation in Azerbaijan, using quarterly data from 2004Q1 to 2024Q4. The analysis is based on out-of-sample forecasts starting from 2019Q1, covering forecast horizons of one, two, four, six and eight quarters ahead. The ML models considered are Random Forest (RF), Extreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), and Elastic Net (ENET). These models are compared against three traditional benchmark models, including the Autoregressive Integrated Moving Average (ARIMA), the Random Walk (RW), and a standard Vector Autoregressive (VAR) model. Forecast accuracy is assessed using the Root Mean Squared Forecast Error (RMSFE) and the Diebold-Mariano (DM) test. The results show that while traditional models perform well at shorter horizons, ML models demonstrate better accuracy as the forecast horizon increases. These findings highlight the potential of ML techniques to be considered by policymakers as a tool for improving medium- to long-term inflation forecasting.

Keywords: Inflation, Forecasting, Forecasting performance, Machine learning
JEL codes: C22, C45, C53, E31, E37

1. Introduction

As price stability is a primary objective of central banks, producing accurate inflation forecasts is crucial for shaping effective monetary policy and achieving inflation targets. In Azerbaijan, inflation is driven by a combination of domestic and external factors. Exchange rate fluctuations, global inflation pressures, prices of agricultural products, and fluctuations in the money supply contribute to consumer inflation. (Rahimov, 2025)

Traditional forecasting methods, such as ARIMA and RW, have been widely used due to their simplicity and established theoretical foundations. These models typically assume linear dynamics and stationarity, which may not fully capture structural breaks, non-linear interactions, or the impact of exogenous shocks. In addition to these univariate models, VAR models are commonly employed in inflation forecasting as they allow for the joint modelling of multiple macroeconomic variables.

Machine learning offers new opportunities for forecasting by handling large datasets, capturing complex economic relationships, and improving prediction accuracy. Unlike traditional models, ML algorithms learn directly from data without relying on pre-specified equations. Although these techniques are gaining popularity in macroeconomic forecasting, their use in Azerbaijan remains limited. This study helps to fill that gap by assessing how well ML models can predict inflation in Azerbaijan, and how they compare to conventional models like ARIMA, RW, and VAR across various forecast horizons.

This research contributes to the broader discussion on the applicability of ML models in economic forecasting by offering empirical evidence from a country where inflation has historically been influenced by supply- and demand-side factors, as well as structural changes. Furthermore, the findings provide insights for central banks and policy institutions on integrating ML-based tools into their forecasting framework.

2. Literature Review

Machine learning has emerged as a valuable tool in economic forecasting, especially for modelling variables characterised by non-linear relationships. In the context of inflation forecasting, several recent studies have shown that ML techniques can outperform traditional econometric models, particularly in environments characterised by complex dynamics or heightened uncertainty.

Mullainathan and Spiess (2017) introduce economists to core ML methods, emphasising their usefulness in prediction tasks and their contrast with traditional econometric approaches. They highlight key ML tools such as regularisation, ensemble methods, decision trees, and neural networks, explaining how they prioritise out-of-sample prediction accuracy over causal inference. The authors argue that ML complements econometrics by offering new ways to handle high-dimensional data and uncover complex patterns without requiring strong model assumptions.

Medeiros and Vasconcelos (2016), using monthly US macroeconomic variables from 1960 to 2011, find that high-dimensional ML models, especially flexible adaptive LASSO, outperform traditional AR and factor models in forecasting macroeconomic variables. Flexible adaptive LASSO consistently yields the lowest forecast errors for both one-step and four-step ahead horizons across most target variables, including inflation. In a subsequent study, Medeiros et al. (2021) extend their US inflation forecasting analysis by employing a set of ML models, including linear shrinkage methods and non-linear algorithms. They demonstrate that ML models, particularly RF, significantly outperform traditional benchmarks, such as RW, AR, and unobserved components models.

Aras and Lisboa (2022) study how ML models, specifically RF and XGBoost, can be used to accurately forecast inflation. Using monthly euro area data, the authors show that tree-based ML models outperform traditional benchmarks such as AR and VAR models in both short- and medium-term forecasts. The study finds that energy prices, labour costs, and expectations are among the most consistently important variables. Overall, the paper demonstrates that explainable ML models offer both high predictive power and transparency, making them valuable tools for central banks and policymakers.

In emerging markets, the literature is more limited but growing. Rodriguez-Vargas (2020) evaluates the performance of various forecasting models, including ARIMA, Bayesian VAR, and ML methods such as RF and XGBoost, in predicting short-term inflation in Costa Rica. The study finds that ML models, particularly XGBoost, outperform traditional models in out-of-sample accuracy, especially during periods of high volatility. The author emphasises the importance of including external variables, such as commodity prices and exchange rates, to improve forecast precision. The results support the use of advanced ML tools by central banks to complement traditional approaches in inflation monitoring and policy decision-making. Araujo and Gaglianone (2023) evaluate the performance of machine learning models, including LASSO, RF and Bagging, for predicting inflation in Brazil using a large set of macroeconomic variables. They find that machine learning models provide more accurate forecasts than traditional models such as autoregressive and factor-based approaches, especially for short-term horizons. The results highlight the importance of flexible model structures and data-driven variable selection in improving inflation forecasts.

To the best of our knowledge, the ML techniques have not yet been applied to inflation forecasting in Azerbaijan. This study aims to fill this gap by exploring the potential of these methods in capturing inflation dynamics. It contributes to the literature by providing empirical evidence on the forecasting performance of machine learning approaches in the context of an oil-exporting economy.

3. Data and Methodology

We use quarterly data for Azerbaijan spanning from 2004Q1 to 2024Q4. The target variable is year-over-year inflation in the Consumer Price Index (CPI). Figure 1 plots the annual average inflation rates of Azerbaijan for the period from 2004 to 2024. As can be seen, Azerbaijan recorded a double-digit inflation rate at the beginning of the sample period, exceeding 20%, primarily driven by a booming economy, expansive government spending, and rapid growth in the money supply. This upward pressure on prices reversed sharply in the second quarter of 2009, when inflation turned negative amid the onset of the Global Financial Crisis and a concurrent drop in global oil prices. As global economic conditions improved, Azerbaijan's inflation rate began to recover and entered a period of relative stability, remaining low and steady for approximately five years. In mid-2014, however, the decline in oil prices triggered a major policy response. The Central Bank of Azerbaijan devalued the national currency in February and December 2015 and introduced a managed floating exchange rate regime. The manat subsequently lost about half of its value against the US dollar. This sharp depreciation led to inflation accelerating at the end of 2015, reaching double-digit levels in 2016. The continued currency weakening into early 2017 kept inflation elevated throughout that year, before it moderated to around 2-3% in 2018 and remained around that level through 2020. A new wave of inflationary pressure emerged in 2021, triggered by a combination of post-pandemic pent-up demand and global supply chain disruptions, including those caused by geopolitical tensions. As a result, inflation surged again, reaching the 13-14% range in 2022. However, beginning in the second half of 2023, inflation began to ease significantly, eventually falling to below 1% in 2024Q2, before accelerating in the second half of the year following the increase in regulated prices.

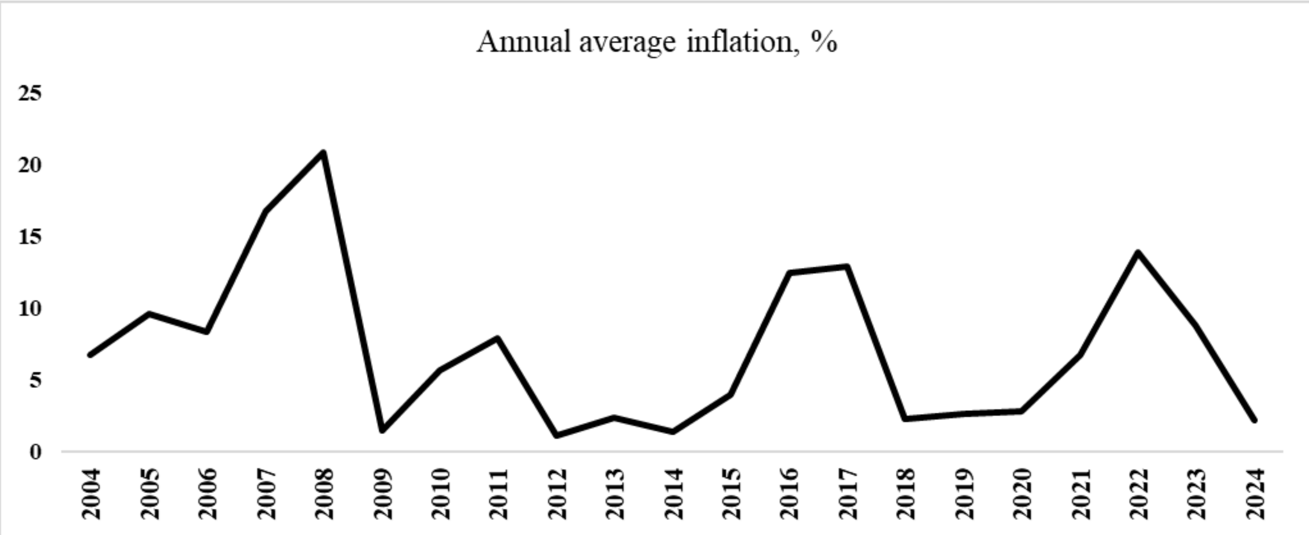


Figure 1. Annual average CPI inflation (2004-2024).

Source: The State Statistical Committee of the Republic of Azerbaijan

The variables used in the ML models include CPI, non-oil real GDP, agricultural producer price index (PPI), manufacturing PPI, food PPI, M2 money supply, non-oil trade-weighted nominal effective exchange rate (NEER), Brent oil price, CPI of main trading partners, and world food price index.

All explanatory variables are transformed into year-on-year percentage changes to align with the dependent variable (inflation), ensure stationarity and eliminate seasonality effects. Augmented Dickey-Fuller unit root test confirms that the transformed series are stationary. We implement two broad categories of forecasting models in this study: ML models and traditional time series benchmarks. The ML group includes RF, XGBoost, SVR and ENET regression. These models are designed to capture complex patterns and interactions among macroeconomic predictors that conventional time series approaches may overlook. While traditional time series models generally require stationary data and assume linear dynamics, ML models are more flexible and do not rely on such assumptions. Nevertheless, transforming the variables helps to improve forecasting performance and facilitates consistency across models.

The RF model, introduced by Breiman (2001), is a non-parametric ensemble learning technique that constructs a multitude of decision trees during training and outputs the average prediction of the individual trees. This averaging helps to reduce model variance and mitigate the risk of overfitting. In our implementation, we optimise key hyperparameters such as the number of trees and maximum tree depth using cross-validation to improve forecast accuracy. The final prediction for a given input x is obtained by averaging the predictions of all individual trees:

$$\hat{y} = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \qquad (1)$$

where B is the total number of trees in the forest, and $T_b(x)$ denotes the prediction from the b-th decision tree.

Extreme Gradient Boosting, also known as XGBoost, is a tree-based ensemble learning algorithm developed by Chen and Guestrin (2016). It is widely recognised for its scalability, computational efficiency, and high predictive accuracy in both regression and classification tasks. It builds an ensemble of weak learners in a sequential manner, where each new tree attempts to correct the residual errors of the previous ones. We fine-tune key hyperparameters, including the learning rate, the number of boosting rounds, and the maximum tree depth, using a grid search combined with time-series cross-validation to avoid data leakage and ensure temporal integrity. The parameter grid includes learning rates in the range {0.01, 0.05, 0.1}, maximum depths from 3 to 10, and boosting rounds from 100 to 1000, with early stopping applied to prevent overfitting. The objective function includes both a differentiable loss function $\ell(y_i, \hat{y}_i)$ and a regularisation term $\Omega(f_t)$, which penalises model complexity:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} \ell\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \qquad (2)$$

Where,

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda\|\omega\|^2 \qquad (3)$$

This regularisation helps control overfitting, particularly in high-dimensional settings. In our implementation, we fine-tune critical hyperparameters such as the learning rate, number of boosting rounds, and maximum tree depth using grid search and cross-validation to achieve optimal forecasting performance.

Support Vector Regression (SVR), proposed by Cortes and Vapnik (1995) and later formalised for regression by Smola and Schölkopf (2004), is a kernel-based machine learning algorithm that estimates a function within a predefined margin of tolerance ε, while minimising model complexity. The model fits a function $f(x) = \omega^T\phi(x) + b$, where $\phi(x)$ maps input data into a higher-dimensional feature space. The objective function of the SVR is expressed as follows:

$$\min_{\omega,b,\xi,\xi^*} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \qquad (4)$$

Subject to

$$y_i - \omega^T\phi(x_i) - b \le \varepsilon + \xi_i, \qquad (5)$$

$$\omega^T\phi(x_i) + b - y_i \le \varepsilon + \xi_i^*, \qquad (6)$$

$$\xi_i, \xi_i^* \ge 0. \qquad (7)$$

We apply the radial basis function kernel to capture non-linear relationships between inflation and its predictors. The hyperparameters C, ϵ and kernel width σ are optimised through grid search with cross-validation.

Elastic Net (ENET) regression, introduced by Zou and Hastie (2005), is a regularised linear modelling technique that combines the strengths of both Least Absolute Shrinkage and Selection Operator (LASSO) and Ridge regression. It addresses the limitations of LASSO, particularly in cases of high multicollinearity or when the number of predictors exceeds the number of observations. ENET applies both L1 and L2 penalties to the regression coefficients, enabling effective variable selection while maintaining grouped variable behaviour. The optimisation problem is defined as:

$$\arg\min_{\beta}\left\{\sum_{i=1}^{n}(y_i - x_i^T\beta)^2 + \lambda\left[\alpha\sum_{j=1}^{p}|\beta_j| + (1-\alpha)\sum_{j=1}^{p}\beta_j^2\right]\right\}$$

where $\lambda \geq 0$ is the overall regularisation parameter and $\alpha \in [0,1]$ controls the balance between LASSO (L1) and Ridge (L2) penalties. Setting $\alpha = 0.5$ gives equal weight to both. We tune λ and α using cross-validation to enhance out-of-sample forecast accuracy.

In contrast, we also estimate three traditional time series benchmarks to provide a reference point for evaluating the performance of the machine learning models. The first is the Autoregressive Integrated Moving Average (ARIMA) model, which is widely used for forecasting univariate time series. An ARIMA(p, d, q) model is expressed as:

$$\Phi(L)(1-L)^d y_t = \Theta(L)\varepsilon_t \qquad (9)$$

Where $\Phi(L) = 1 - \phi_1 L - \cdots - \phi_1 L^p$ and $\Theta(L) = 1 + \theta_1 L + \cdots + \theta_1 L^q$ are lag polynomials, L is the lag operator, and ε_t is a white noise error term.

The second benchmark is the Random Walk (RW) model, a robust yet straightforward forecasting method that assumes inflation follows a stochastic process with no mean reversion. The RW model is a special case of ARIMA($0,1,0$), defined as:

$$y_t = y_{t-1} + \varepsilon_t \qquad (10)$$

Under this approach, future values are predicted based solely on the most recent observation. Despite its simplicity, the RW often performs competitively in macroeconomic forecasting, making it a strong baseline for evaluating predictive improvements.

As a third benchmark, we include the vector autoregression (VAR) model, initially proposed by Sims (1980), which captures the dynamic interrelationships among multiple endogenous macroeconomic variables without requiring strong theoretical priors on variable ordering. Following Rahimov (2025), the VAR model includes six variables: CPI, agricultural PPI, M2 money supply, non-oil real GDP, CPI of main trading partners and non-oil trade-weighted NEER, all expressed as year-over-year percentage changes. A VAR(p) model for a K-dimensional vector of variables y_t is specified as:

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \varepsilon_t \qquad (11)$$

Where A_i are $K \times K$ coefficient matrices, and $\varepsilon_t \sim N(0, \Sigma)$ is a vector of innovations with a constant covariance matrix Σ. Each equation in the VAR is estimated using ordinary least squares, exploiting the fact that all regressors are predetermined under the assumption of no contemporaneous feedback.

By comparing the results from these diverse modelling strategies, we aim to assess whether ML techniques provide meaningful improvements in forecast accuracy over traditional time series approaches in the context of inflation forecasting.

4. Forecast Design and Evaluation

The expanding window approach is applied in this study, where the training sample grows recursively by including one additional observation at each step. This method reflects realistic forecasting settings in which models are re-estimated as new data becomes available, enabling better adaptation to evolving economic conditions. All forecasts are generated dynamically, meaning that each prediction beyond the initial period relies on model estimates rather than observed future values, thereby ensuring a fair and consistent comparison across models.

The training period spans from 2004Q1 to 2018Q4. Forecasts are generated recursively for each horizon beginning in 2019Q1 and ending in 2024Q4. We produce forecasts for one-, two-, four-, six- and eight-quarter horizons. The expanding window strategy simulates real-world forecast conditions, where the model is updated each period with newly available data.

To assess the forecasting accuracy of the models, we employ the root mean squared forecast error (RMSFE). For a given variable i over a forecast horizon of h periods, the RMSFE is calculated as:

$$RMSFE_i = \sqrt{\left(\frac{1}{h}\right)\sum_{i=1}^{h}(y_{T+i} - \tilde{y}_{T+i})^2}, \quad (12)$$

where y_{T+i} is the actual value, while $\tilde{y}_{T+i}$ is the forecasted value.

In this study, RMSFE is reported in relative terms. Specifically, it is computed as the ratio of the RMSFE from the ML models to that of a benchmark model. A relative RMSFE below one indicates superior forecasting performance of the ML model, while a value above one suggests inferior performance.

While some studies suggest that averaging forecasts across multiple models can enhance predictive accuracy by mitigating individual model weaknesses, this approach is not pursued in the present analysis. Instead, the focus is on evaluating the standalone performance of each model to highlight its relative strengths across forecast horizons.

As our focus is on evaluating ML models, we select the RW, ARIMA and VAR models as benchmarks. To determine whether the differences in forecast accuracy are statistically significant, we apply the Diebold-Mariano (DM) test (1995). The test is based on the loss differential between the forecast errors of two models, defined as:

$$d_\tau = g\left(e_\tau^{(2)}\right) - g\left(e_\tau^{(1)}\right), \quad (13)$$

where $e_\tau^{(i)}$ is the forecast error from model i, and $g(\cdot)$ is a general loss function, typically the squared error.

The DM test statistic is calculated as:

$$DM = \frac{\bar{d}}{\sqrt{\frac{Var(d_T)}{n}}} \quad (14)$$

Where $\bar{d}$ is the mean of the loss differentials d_τ, $Var(d_\tau)$ is their variance, and n is the number of forecast observations. The null hypothesis states that there is no statistically significant difference in forecast accuracy between the model of interest and the benchmark.

5. Results and Discussion

This section evaluates the forecasting performance of ML models relative to three benchmark models. Two primary metrics are used to assess forecast accuracy. The first is the RMSFE[1], which indicates the magnitude of forecast errors produced by each ML model relative to the benchmarks. The second is the DM test, which provides a formal statistical comparison of predictive accuracy. A p-value below 0.05 is interpreted as evidence that the ML model delivers significantly better forecasts than the benchmark.

Table 1. Relative RMSFE of Machine Learning Models vs Traditional Benchmark Models

Horizon	Model	ML vs ARIMA	ML vs RW	ML vs VAR
1-step	RF	1.22	**0.87**	**0.81**
	XGB	1.36	**0.97**	**0.90**
	SVR	1.65	1.17	1.09
	ENET	1.68	1.20	1.12
2-step	RF	**0.73**	**0.54**	**0.69**

[1] We have also calculated the relative Mean Absolute Errors between the ML and benchmark models. The results are similar to RMSFEs, which can be shared upon request.

Horizon	Model	ML vs ARIMA	ML vs RW	ML vs VAR
	XGB	**0.69**	**0.51**	**0.65**
	SVR	1.12	**0.83**	1.06
	ENET	1.09	**0.81**	1.03
4-step	RF	**0.47**	**0.35**	**0.44**
	XGB	**0.48**	**0.36**	**0.44**
	SVR	**0.75**	**0.56**	**0.69**
	ENET	**0.78**	**0.58**	**0.72**
6-step	RF	**0.34**	**0.26**	**0.29**
	XGB	**0.34**	**0.27**	**0.29**
	SVR	**0.62**	**0.49**	**0.53**
	ENET	**0.67**	**0.53**	**0.57**
8-step	RF	**0.27**	**0.24**	**0.24**
	XGB	**0.25**	**0.22**	**0.22**
	SVR	**0.56**	**0.49**	**0.49**
	ENET	**0.70**	**0.62**	**0.61**

Table 1 reports the relative RMSFE values for the ML models across different forecast horizons. At the shortest horizon, corresponding to one-quarter-ahead forecasts, some ML models perform better than certain benchmarks. For instance, RF and XGB yield relative RMSFEs of less than one when compared to the RW and VAR models. However, SVR and ENET still produce forecast errors larger than all three benchmarks. This suggests that while traditional time series models often remain competitive at short horizons, specific ML models like RF and XGB can offer improved accuracy even in the near term.

As the forecast horizon extends, the relative performance of the ML models improves noticeably. By the two-step horizon, ML models begin to deliver comparable or slightly better results than the benchmarks, particularly when compared to the RW. However, the most significant improvements emerge from the four-step horizon onwards. At this point, ML models consistently produce RMSFE values below one, which indicates better predictive accuracy.

At the six- and eight-step horizons, the differences become substantial. RF and XGB achieve the lowest forecast errors among all models. Their performance is robust across all benchmark comparisons. These results reflect the ability of ML algorithms to capture underlying non-linear patterns and long-term dependencies that may be difficult to model using linear time series techniques.

Table 2. Diebold-Mariano Test p-values (Machine Learning vs Benchmark Models)

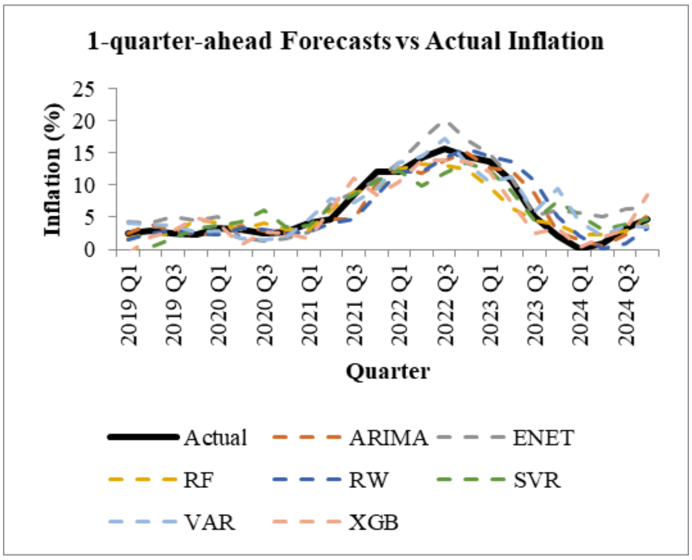
Horizon	Model	ML vs ARIMA	ML vs RW	ML vs VAR
1-step	RF	0.44	0.56	0.48
	XGB	0.10	0.86	0.73
	SVR	0.08	0.50	0.56
	ENET	0.05	0.43	0.57
2-step	RF	0.34	0.15	0.30
	XGB	0.24	0.14	0.31
	SVR	0.70	0.54	0.71
	ENET	0.79	0.54	0.91
4-step	RF	0.08	0.07	0.06
	XGB	0.07	0.09	0.09
	SVR	0.31	0.10	0.11
	ENET	0.35	0.13	0.19
6-step	RF	0.00	0.01	0.00
	XGB	0.01	0.02	0.00
	SVR	0.18	0.02	0.00
	ENET	0.08	0.00	0.01
8-step	RF	0.00	0.00	0.00
	XGB	0.00	0.01	0.01
	SVR	0.02	0.00	0.00
	ENET	0.11	0.00	0.00

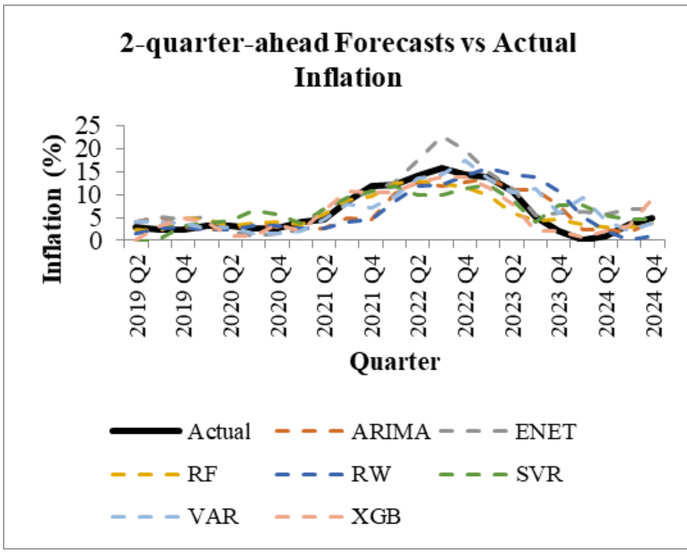
We present the results of the Diebold-Mariano tests in Table 2. At the one-step horizon, the SVR and ENET models perform significantly worse than the ARIMA benchmark at 10% and 5% significance levels, respectively, evidenced by p-values. This finding aligns with the relative RMSFE values reported in Table 1, where SVR and ENET have ratios above one, indicating worse performance.

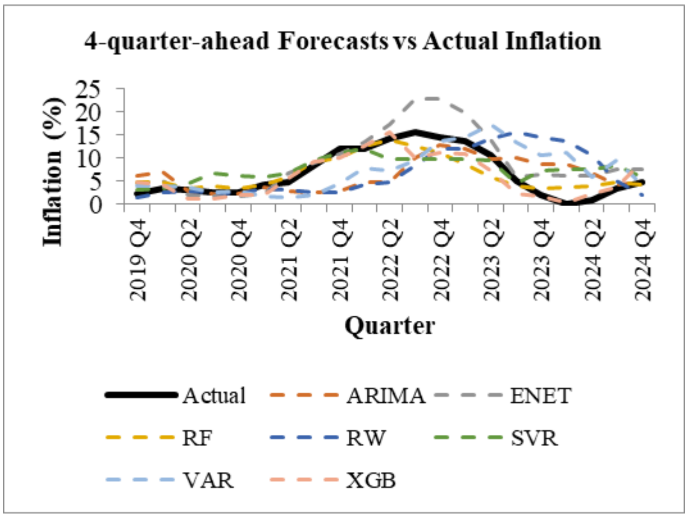
At the two-step horizon, the DM test results show no statistically significant differences in forecast accuracy between the ML models and the benchmark models. All p-values exceed conventional significance thresholds of

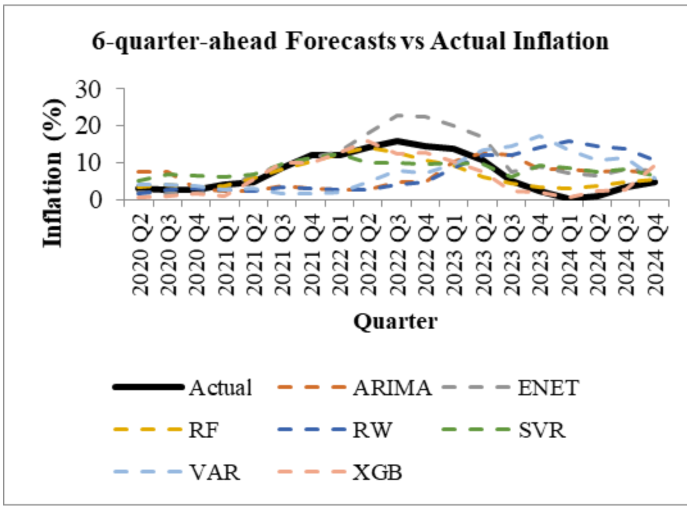
5%, indicating that the observed performance differences are not statistically robust. At the four-step horizon, all ML models outperform the benchmark models in terms of RMSFE. However, none of these differences are statistically significant, indicating that the improvements are not robust. Overall, the evidence at these horizons does not support a consistent forecasting advantage for ML models at 5% significance level.

The most convincing evidence in favour of ML models emerges at the six and eight-step horizons. At these longer forecast horizons, RF and XGB exhibit highly statistically significant improvements across all benchmark comparisons. SVR and ENET also perform well, with several p-values below 0.05, particularly in comparison to the RW and VAR models. At the eight-step horizon, the evidence is particularly strong: RF, XGB and SVR show statistically significant improvements over ARIMA, RW and VAR. ENET also delivers significant gains over RW and VAR, though not ARIMA.









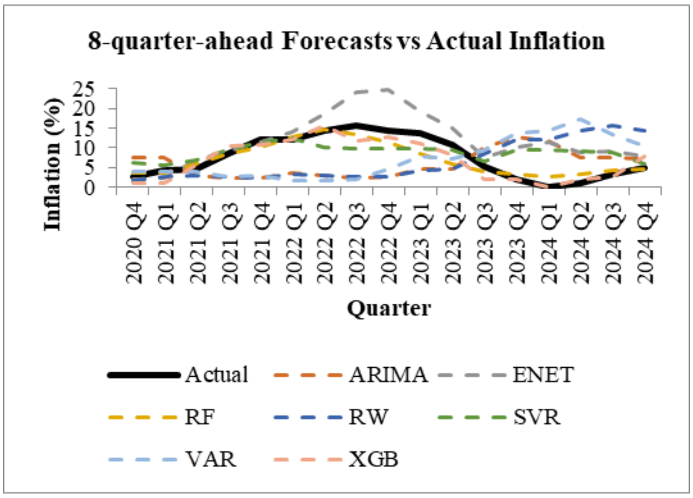


Figure 2. Forecasts vs Actual Inflation

Furthermore, Figure 2, which illustrates actual inflation against forecasts from each model, visually confirms the patterns observed in the tables. At the shortest horizons, ARIMA tend to track actual inflation more closely than the ML models. As the forecast horizon lengthens, machine learning models, especially RF and XGB, produce smoother and more aligned forecasts that better capture turning points and underlying volatility in inflation. This reflects the ML models' ability to learn long-term dependencies and non-linear relationships in the data that traditional linear models may fail to represent.

The comparative evaluation based on Tables 1 and 2, along with the visual illustration in Figure 2, reveals a clear pattern of forecast performance that varies by horizon. In the short term, the ML models either underperform or do not significantly outperform traditional time series methods in terms of predictive accuracy. This may be attributed to the strong autoregressive structure of inflation in the near term and the tendency of ML models to require more data to effectively detect subtle patterns. However, as the forecast horizon increases, the ML models begin to outperform their traditional counterparts. These improvements are evident not only in relative forecast error metrics but also in statistically significant differences at longer horizons.

The differences in model performance across forecast horizons largely stem from how each model processes information. Traditional models, such as ARIMA, RW, and VAR, tend to perform well at shorter horizons. ARIMA and RW rely heavily on recent values of the target variable, capturing strong short-term autocorrelation commonly seen in inflation. VAR models, while incorporating multiple macroeconomic indicators, are still limited by their linear structure and fixed lag specification, which can constrain their flexibility in adapting to changing dynamics. In contrast, as the forecast horizon extends, the strengths of ML models, especially RF and XGB, become increasingly evident. These models are better at picking up complex, non-linear patterns and interactions between variables, which often take more time to materialise. Their advantage at longer horizons reflects this ability to recognise deeper structures in the data, such as delayed effects from macroeconomic indicators, that traditional models may miss.

6. Conclusion

This study highlights that while traditional series models remain competitive at shorter horizons, ML models become increasingly effective as the forecast horizon extends. Using quarterly data on Azerbaijan's inflation and related domestic and external variables from 2004Q1 to 2024Q4, the models were trained on data up to 2018Q4, with forecasts evaluated out-of-sample from 2019Q1 onward. The analysis covers forecast horizons of one, two, four, six, and eight quarters ahead. The ML models considered are RF, XGBoost, SVR, and ENET. These are compared against three traditional benchmark models, including the ARIMA, the RW, and a standard VAR model. Forecast accuracy is assessed using the RMSFE and the DM test. The results indicate that ML models offer more accurate forecasts at longer horizons, particularly beyond four quarters, by better capturing turning points and long-term patterns.

These findings have important implications for macroeconomic forecasting and policy analysis. In countries like Azerbaijan, where inflation dynamics are influenced by a combination of demand- and supply-side factors, structural changes, and external shocks, traditional linear models may sometimes struggle to capture the full complexity of these processes. ML approaches enhance forecasting performance by identifying complex patterns and responding to structural shifts that conventional models often miss. For central banks seeking reliable medium- to long-term inflation forecasts to inform monetary policy, ML models present a promising complement to existing frameworks.

References

Aras, S. and Lisboa, P.J., 2022. Explainable inflation forecasts by machine learning models. *Expert Systems with Applications*, *207*, p.117982.

Araujo, G.S. and Gaglianone, W.P., 2023. Machine learning methods for inflation forecasting in Brazil: New contenders versus classical models. *Latin American Journal of Central Banking*, *4*(2), p.100087.

Breiman, L., 2001. Random forests. *Machine learning*, *45*, pp.5-32.

Chen, T. and Guestrin, C., 2016, August. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

Cortes, C. and Vapnik, V., 1995. Support-vector networks. *Machine learning*, *20*, pp.273-297.

Diebold, F.X. and Mariano, R.S., 1995. Com paring predictive accuracy. *Journal of Business and Economic Statistics*, *13*(3), pp.253-263.

Medeiros, M.C. and Vasconcelos, G.F., 2016. Forecasting macroeconomic variables in data-rich environments. *Economics Letters*, *138*, pp.50-52.

Medeiros, M.C., Vasconcelos, G.F., Veiga, Á. and 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.

Mullainathan, S. and Spiess, J., 2017. Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, *31*(2), pp.87-106.

Rahimov, V., 2025. "A Bayesian VAR Model for Inflation: The Case of Azerbaijan. *Financial and Credit Activity Problems of Theory and Practice*, 4(63), pp. 180–191.

Rodríguez-Vargas, A., 2020. Forecasting Costa Rican inflation with machine learning methods. *Latin American Journal of Central Banking*, *1*(1-4), 100012.

Sims, C.A., 1980. Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, pp.1-48.

Smola, A.J. and Schölkopf, B., 2004. A tutorial on support vector regression. *Statistics and computing*, *14*, pp.199-222.

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