



Research article

AI meets economics: Can deep learning surpass machine learning and traditional statistical models in inflation time series forecasting?

Ezekiel NN Nortey¹, Edmund F. Agyemang^{1,2}, Enoch Sakyi-Yeboah^{1*}, Obu-Amoah Ampomah³, Louis Agyekum¹

¹ Department of Statistics and Actuarial Science, College of Basic and Applied Sciences, University of Ghana, Ghana

² School of Mathematical and Statistical Science, College of Sciences, University of Texas Rio Grande Valley, USA

³ Department of Statistics, Western Michigan University, Kalamazoo-USA.

***Correspondence:** Email: esyeboah@ug.edu.gh.

Abstract: This study examined the forecasting ability of deep learning (DL) and machine learning (ML) models against benchmark traditional statistical models for the monthly inflation rates in the USA. The study compared various DL and ML models like transformers, linear regression, gradient boosting (GB), extreme gradient boosting (XGBoost), and adaptive boosting (AdaBoost) with traditional baseline time-series models like autoregressive integrated moving averages (ARIMA) and exponential smoothing (ETS) with Holt-Winters seasonal method utilizing data sourced from the Federal Reserve Bank of St. Louis. The study consistently showed that all DL and ML models outperformed the traditional approaches. In particular, the Transformer (RMSE = 0.0291, MAE = 0.0221) exhibited the lowest error rates, suggesting high forecasting accuracy for the monthly inflation data. In contrast, ARIMA (RMSE = 0.2038, MAE = 0.1895) and ETS (RMSE = 0.1619, MAE = 0.1455) models displayed higher error rates. Due to the ability of DL and ML models to process large volumes of data and identify underlying patterns, they are particularly effective for dynamic and evolving datasets. This study explored how novel DL modeling techniques could augment the accuracy and reliability of economic forecasting. The results indicate that policymakers and economists could leverage DL and ML models to gain deeper insights into inflation dynamics and implement more informed strategies to combat inflationary pressure. Hybrid methods should be further investigated in future studies to enhance economic forecasting accuracy.

Keywords: transformer; gradient boosting; extreme gradient boosting; adaptive boosting; linear regression; inflation

JEL Codes: G11, G12, G17, C22

1. Introduction

The rapid development of artificial intelligence (AI) has begun a transformative wave that covers a broad spectrum of domains, from finance, economics, and manufacturing to health, weather, and similar fields. These domains rely substantially on precise forecasts to predict future demands and sales. In finance and economics, forecasting assists organizations in managing inventory optimization as well as improving the performance of the supply chain. This is crucial for decision-making in areas such as financial management, marketing strategies, and risk mitigation (Masini et al., 2023). In manufacturing and health, accurate predictions protect against over- or under-production by maintaining appropriate stock levels. Not only does this streamline business operations but it also reduces operational costs and increases overall operational efficiency (Vaughan et al., 2023; Bonci et al., 2024). Moreover, accurate predictions aid in planning and reducing the risk in weather and road traffic accident studies (Agyemang et al., 2023; Bouallègue et al., 2024). Market fluctuations are an integral part of any business and organization, which is why they must rely on accurate forecasting to manage effectively and maintain a competitive edge. Many researchers prefer to employ standard models like ARIMA, SARIMA, and ETS for time series analysis. These models are preferred for their simplicity and for being less sensitive to data assumptions and theoretically well-grounded. Consequently, they are well-suited for future predictions (Valderrama et al., 2024). However, with the advancement of DL and ML models, there has been a significant rise in the use of these techniques for predicting time series data.

ML techniques such as bagging and boosting have attracted a lot of attention, as they can identify intricate patterns and hidden relationships in time series data (Masini et al., 2023). ML approaches are unique in the sense that they learn from raw data, enabling them to process large volumes of data in a more efficient manner. Furthermore, ML models can adapt and enhance their predictions as time progresses and are ideally suited for dynamic and evolving datasets (Arora et al., 2024). They have the capacity to process and analyze large datasets, providing highly accurate predictions, and often outperforming traditional methods (Westergaard et al., 2024). Although classical approaches such as ARIMA, ETS, and STL are still useful for their high usability and theoretical grounding, the emergence of ML methods is changing the landscape of time series forecasting. The focus of this study is to determine whether ML and DL models such as transformers, linear regression, GB, XGBoost, and AdaBoost can provide more accurate and reliable inflation forecasts compared to well-established traditional models like ARIMA and ETS.

The domain of economics is continuously changing, and with the advent of ML models, there is a growing tendency for ML models to outperform conventional traditional statistical models for inflation time-series forecasting. ML models can enhance accuracy when predicting inflation, enabling better cross-sectoral and policy-making awareness in the economic community (Zhang et al., 2021). Despite the promising results of ML models in the literature, there is still a lot of research to be done

in assessing whether ML models yield superior performance in inflation time-series forecasting when compared with benchmark traditional models (Adnan et al., 2021). AI models have the potential to revolutionize inflation time-series forecasting, and this potential has sparked more interest and investment in AI-powered tools and research in the economics domain. For decades, traditional statistical models such as ARIMA have been the cornerstone of time series forecasting due to their grounded mathematical foundation and simplicity (Kontopoulou et al., 2023). Aras and Lisboa (2022) discussed the application of ML techniques, especially random forest and other tree-based methods for enhancing inflation predictions. It was observed that ML methods have an advantage over traditional models: they can accommodate many predictors and explore high-order interactions within the data. Results also show that ML models not only outperform traditional models in terms of prediction accuracy but also offer enhanced interpretability, providing economists with insight to better understand the underlying factors driving inflation.

The advent of DL models has however introduced a paradigm shift, with researchers exploring their potential to outperform classical statistical approaches as well as ML methods (Sezer et al., 2020). Interestingly, while ML algorithms generally display better prediction performance in various applications, including finance and health, the superiority of these methods over DL models in inflation forecasting is not conclusively established. Neural networks, particularly those with memory capabilities like LSTM and GRU, are consistently among the top-performing models, indicating their relevance in capturing complex temporal dynamics (Kontopoulou et al., 2023; Gunnarsson et al., 2024). A detailed analysis by Paranhos (2025) focused on the application of DL models for inflation forecasting. The study compared the performance of DL and ML with traditional models. The results showed that DL models capturing long-term dependencies and complex temporal patterns in the data outperformed ML and traditional models, making them particularly suitable for forecasting tasks that involve intricate and nonlinear relationships.

Another study by Makala and Li (2021) demonstrated that ML models, specifically SVM, can surpass traditional statistical models such as ARIMA in time series forecasting. The research, which focused on predicting gold prices, revealed that SVM significantly outperformed ARIMA in terms of accuracy. Utilizing performance metrics like RMSE and MAPE, the SVM model achieved notably lower error rates (RMSE of 0.028 and MAPE of 2.50) compared to the ARIMA model (RMSE of 36.18 and MAPE of 2897). This suggests that SVM can handle complex, nonlinear data patterns and is a more effective tool for time series forecasting, highlighting the potential of machine learning to enhance predictive analytics in various fields. Despite the promise of ML and DL models, their “black-box” nature and the scarcity of explainable AI (XAI) applications in the literature are notable concerns, highlighting the need for transparency and interpretability in economic forecasting (He et al., 2019).

This study assesses the predictive power of advanced DL and ML techniques against standard traditional baseline models in forecasting monthly inflation rates in the USA. The remainder of the paper is organized as follows: Section 2 discusses the data and methods used for the study. Section 3 provides the mathematical framework of the study. Section 4 discusses the results of the study and provides a comparison between the RMSE and MAE obtained here with that of other studies in the literature that used economics datasets. Section 5 concludes the study and provides recommendations for further work.

2. Data and method of analysis

The USA 10-Year Breakeven Inflation Rate (T10YIEM) data used in this study was retrieved from the website of the Federal Reserve Bank of St. Louis and is openly available and can be assessed at the repository via <https://fred.stlouisfed.org/series/T10YIEM>. To facilitate time-series analysis, the *month* column was converted to a date-time format using the `pd.to_datetime()` function. This conversion ensured that the data was correctly interpreted as time-series data. The *month* column was set as the index, and the data was resampled to a monthly frequency using the mean values. This resampling was performed using the `resample('MS').mean()` method, which helped standardize the time intervals and smooth out any irregularities in the data.

The dataset was split into training and test sets, with 245 months used for training and the remaining 12 months reserved for testing the performance of the models. These included both traditional time series models such as ARIMA and ETS with the Holt-Winter seasonal method. Likewise, ML and DL models, which included transformers, linear, GB, XG Boost, and AdaBoost regressions, were employed due to their high predictive accuracy and ability to handle complex patterns in data. The best ARIMA was identified by considering seasonality with a period of 12 months. The ML Forecast library was used to handle time series forecasting with machine learning models. This was configured with various lag features and transformations, including rolling means and expanding means. Target transformations using differences were also applied to stabilize the time series. The features for modeling the ML models were created from the time series data itself using lags of the target variable at different optimal time intervals (1, 3, 5, 7, and 12 months).

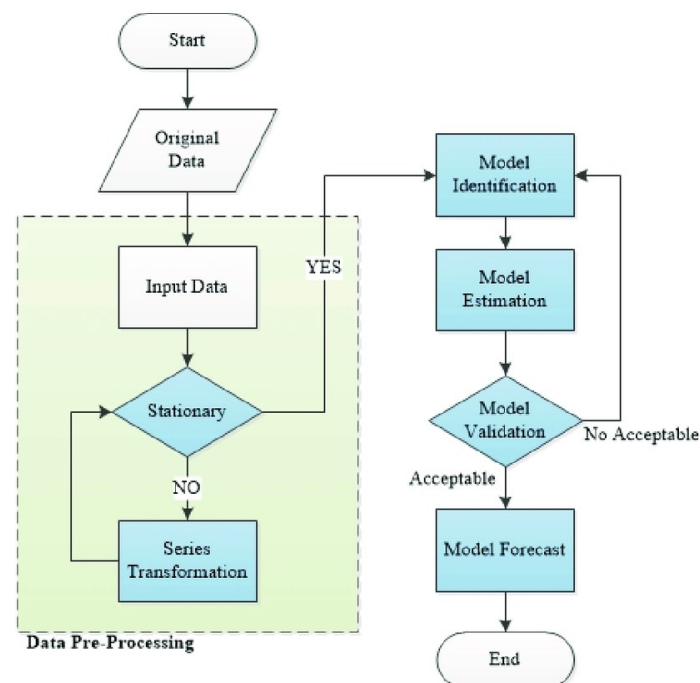


Figure 1. Workflow adopted for ARIMA and ETS.

Transformations of these lags such as rolling and expanding means were applied to capture different aspects of the time series. Each model was then fitted on the training data, and forecasts were

generated for the test set. The actual inflation rates were then plotted against the predicted values from each model to visually assess the model's performance. Prominent time series model evaluation metrics such as MAE, RMSE, MAPE, and SMAPE were used to assess the model's performance. Forecasts from June 2024 until the end of December 2025 were then generated using both traditional and ML models. These future forecasts provided insights into expected trends in the USA's monthly inflation rates, aiding in decision-making and strategic planning.

The workflow adopted for the traditional baseline ARIMA and ETS models is depicted in Figure 1.

3. Mathematical framework of the study

3.1. ARIMA model for time-series forecasting formulation

ARIMA is a popular statistical method used for analyzing and forecasting time-series data. It combines three key components: autoregressive (AR), integrated (I), and moving average (MA). Each component plays a critical role in capturing the underlying patterns in the data.

3.1.1. Autoregressive (AR) component

The AR part of ARIMA models the variable of interest as a linear combination of its past values. Mathematically, an AR model of order p (AR(p)) is expressed in Equation (1) as:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t \quad (1)$$

where y_t is the value at time t , $\phi_1, \phi_2, \dots, \phi_p$ are the parameters of the model, and ϵ_t is the white noise error term.

3.1.2. Integrated (I) component

The I part of ARIMA involves differencing the time series to achieve stationarity. A time series is said to be stationary if its statistical properties such as mean and variance are constant over time. The order of differencing required to make the series stationary is denoted by d . The differenced series y'_t can be expressed in Equation (2) as:

$$y'_t = y_t - y_{t-1} \quad (2)$$

If differencing is applied d times, the model becomes ARIMA (p, d, q).

3.1.3. Moving average (MA) component

The MA part models the error term as a linear combination of past error terms. An MA model of order q (MA(q)) is expressed in Equation (3) as:

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} \quad (3)$$

where $\theta_1, \theta_2, \dots, \theta_q$ are the parameters of the model.

3.1.4. ARIMA model

Combining these three components, the ARIMA model of order (p, d, q) is formulated in Equation (4) by:

$$y'_t = \phi_1 y'_{t-1} + \phi_2 y'_{t-2} + \cdots + \phi_p y'_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \quad (4)$$

where y'_t is the differenced series.

3.2. ETS for time-series forecasting

ETS is a versatile and widely used method for forecasting time-series data, particularly effective in capturing trends and seasonality. ETS models consist of three main components: error (E), trend (T), and seasonality (S). The study employed the Holt-Winters seasonal method.

3.2.1. Holt-Winters seasonal method

The Holt-Winters method incorporates seasonality along with level and trend. It can be either additive or multiplicative. For the additive model considered in this study, the forecast equations for the level l_t , seasonal s_t , and trend b_t is given by Equations (5) – (8):

$$\hat{y}_{t+h} = l_t + hb_t + s_{t+h-m(k+1)} \quad (5)$$

$$l_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (6)$$

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad (7)$$

$$s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m} \quad (8)$$

where α is the level smoothing parameter ($0 < \alpha < 1$), β is the trend smoothing parameter ($0 < \beta < 1$), γ is the seasonal smoothing parameter ($0 < \gamma < 1$), h is the forecast horizon, m is the number of periods in a season, s_t is the seasonal component, k is the integer part of $(\frac{h-1}{m})$, (y_t) is the actual value at time t , and \hat{y}_t is the forecasted value at time t . By adjusting the level, trend, and seasonal components, ETS models can provide accurate and adaptable forecasts.

3.3. Linear regression for time-series forecasting of monthly inflation

Linear regression is a fundamental supervised machine learning algorithm used for modeling the relationship between a dependent variable and one or more explanatory variables. The linear regression model used in the study is expressed by (9):

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_n x_{n,t} + \epsilon_t \quad (9)$$

where y_t is the predicted value of the dependent variable at time t , β_0 is the y-intercept, $\beta_1, \beta_2, \dots, \beta_n$ are the coefficients representing the change in y_t for a unit change in the corresponding $x_{i,t}$; $x_{1,t}, x_{2,t}, \dots, x_{n,t}$ where $x_{1,t}, x_{2,t}, \dots, x_{n,t}$ are the explanatory variables at time t , and ϵ_t is the error term. The

parameters $\beta_0, \beta_1, \dots, \beta_n$ were estimated using the method of least squares, which minimizes the sum of the squared differences between the observed values y_t and the predicted values \hat{y}_t as given in Equation (10) by:

$$\min \sum_{t=1}^T (y_t - \hat{y}_t)^2 \quad (10)$$

where T is the total number of observations. For forecasting monthly inflation, the study incorporated lagged inflation rates as explanatory variables of the linear regression as indicated in section 2 of the study.

3.4. Gradient boosting (GB) regression

Gradient boosting regression sequentially builds an ensemble of trees by combining multiple weak learners to create a strong predictive model, where each tree corrects the errors of the previous ones. The GB model is formulated by (11) as:

$$F_m(x) = F_{m-1}(x) + \eta h_m(x) \quad (11)$$

where $F_m(x)$ is the ensemble model after m iterations, η is the learning rate, and $h_m(x)$ is the m -th weak learner. The weak learner $h_m(x)$ is trained to minimize the loss function L given by (12):

$$h_m = \arg \min_h \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + h(x_i)) \quad (12)$$

where y_i are the actual values, x_i are the feature vectors, and n is the number of observations.

3.5. Extreme gradient boosting (XGBoost) regression

XGBoost is an advanced implementation of GB designed for speed and performance. It introduces regularization to prevent overfitting and uses a more efficient computation method. The model is given by (13):

$$F_m(x) = \sum_{k=1}^m h_k(x) \quad (13)$$

The objective function to be minimized is given by (14):

$$\mathcal{L}(\theta) = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_{k=1}^m \Omega(h_k) \quad (14)$$

where $\Omega(h_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2$ is the regularization term, γ and λ are regularization parameters, T is the number of leaves in the tree, and w_j are the leaf weights.

3.6. Adaptive boosting (AdaBoost) regression

AdaBoost works by adjusting the weights of training instances so that harder cases are given more focus in subsequent iterations. The model is expressed in (15) as:

$$F_m(x) = \sum_{k=1}^m \alpha_k h_k(x) \quad (15)$$

where α_k is the weight assigned to the k -th weak learner, and $h_k(x)$ is the k -th weak learner. The weights α_k are computed based on the error of the k -th learner given in (16) by:

$$\alpha_k = \frac{1}{2} \ln \left(\frac{1-e_k}{e_k} \right) \quad (16)$$

where e_k is the weighted error rate of the k -th learner.

3.7. Evaluation metrics

To evaluate the performance of the models used in this study, several error metrics were calculated.

(a) Mean absolute error (MAE) is the average of the absolute differences between the predicted and actual values. It is calculated in (17) as:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (17)$$

(b) Root mean square error (RMSE) measures the square root of the average of the squared differences between the predicted and actual values. It is given by (18) as:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (18)$$

(c) Mean absolute percentage error (MAPE) measures the average of the absolute percentage differences between the predicted and actual values. It is computed in (19) as:

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (19)$$

(d) Symmetric mean absolute percentage error (SMAPE) is a measure of accuracy based on relative percentage errors. It is calculated by (20) as:

$$\text{SMAPE} = \frac{100\%}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{\frac{|y_i| + |\hat{y}_i|}{2}} \quad (20)$$

The interpretation of typical MAPE values adopted from Yıldırım et al. (2019) employed in the study is given in Table 1.

Table 1. MAPE evaluation criteria.

MAPE, %	Evaluation
$\text{MAPE} \leq 10\%$	High accuracy forecasting
$10\% < \text{MAPE} \leq 20\%$	Good forecasting
$20\% < \text{MAPE} \leq 50\%$	Reasonable forecasting
$\text{MAPE} > 50\%$	Inaccurate forecasting

4. Results and findings

In this section, we present the results and discuss the findings of the study.

Table 2. Descriptive statistics of monthly inflation.

Statistic	Minimum	Mean	25%	Median	75%	Maximum
Value (%)	0.25	2.09	1.83	2.19	2.36	2.88

Table 2 provides the descriptive statistics for the monthly inflation in the USA, highlighting key summary measures. The minimum value of inflation recorded is 0.25%, while the maximum value reaches 2.88%. The mean inflation rate is 2.09%, indicating the average level of inflation over the observed period. The median value, at 2.19%, shows the midpoint of the inflation rates. The 25th percentile (1.83%) and the 75th percentile (2.36%) provide insights into the spread and distribution of the data (with an interquartile range of 0.53%), showing that the middle 50% of the inflation rates fall within this range.

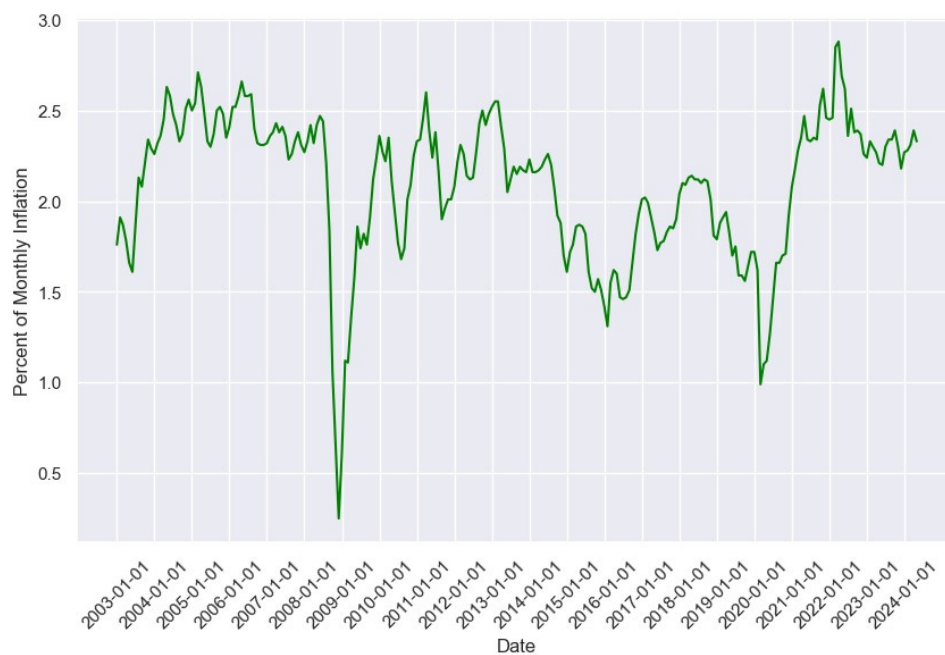
**Figure 2.** Time-series plot of monthly inflation over time.

Figure 2 presents the time series plot of percentage monthly inflation over time in the USA from January 2003 to May 2024. Throughout this period, the inflation rate exhibited significant fluctuations. Initially, from 2003 to 2008, there was a general upward trend, peaking at around 2.70%. However, this was followed by a sharp decline around 2009, likely coinciding with the global financial crisis, where inflation rates plummeted to approximately 0.25%. The drop in inflation during the 2008 economic downturn in the USA was due to several factors. One of the main reasons was a significant decline in energy prices, which led to a reduction in consumer prices. Additionally, weak economic activity and low commodity prices, including crude oil, created downward pressure on prices, resulting in the slowest 12-month gain in consumer prices since 1954 (Reed, 2014).

Despite the economic slump and high unemployment, core inflation only fell by one percentage point, from 2.20% in 2007 to 1.20% in 2009, a situation some economists refer to as the “missing deflation” puzzle. Post-financial crisis, the inflation rate worsened and fluctuated between 1.50% and

2.60% up until around 2017. A noticeable dip is observed again around 2020, possibly reflecting economic impacts due to the COVID-19 pandemic, before the rate rises sharply again. Figure 2 demonstrates the volatility and cyclic nature of the inflation rate over the years, with external economic events significantly impacting the trends. The period from 2021 onward shows a steep rise in inflation, peaking near 3%, followed by another fluctuating trend until May 2024. Figure 2 indicates that USA inflation rates were highly volatile from 2003 to 2024, reflecting the economy's sensitivity to major events like the 2008 financial crisis and the 2020 COVID-19 pandemic. Sharp declines during these periods highlight the significant impact of economic disruptions, while subsequent recoveries suggest the effects of policy interventions.

4.1. Trend and seasonality analysis

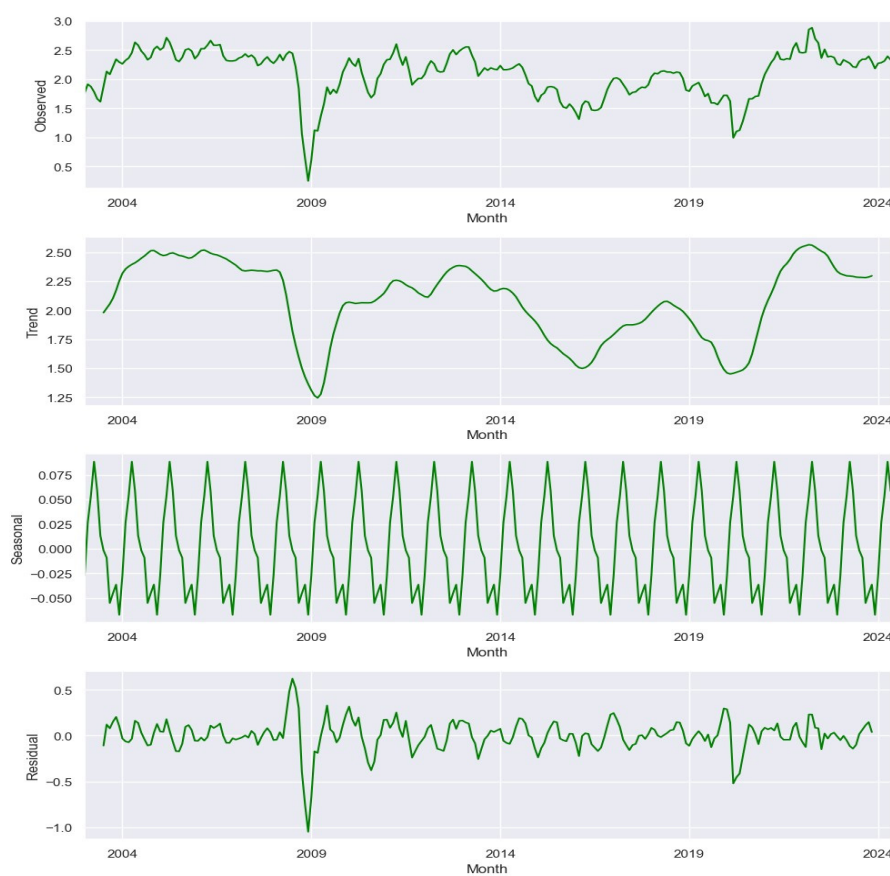


Figure 3. Seasonal decomposition of monthly inflation into trend, seasonality, and residuals.

The study employed a time-series decomposition method to decompose the USA inflation time series into its trend, seasonal, and residual (irregular) components, as evident in Figure 3. The trend component depicts the long-term movement of inflation rates over the observed period and captures the overall direction and structural changes in the inflation rates. We observed a general increase in inflation until the 2008 financial crisis, followed by a decline and subsequent increment. The trend also highlights periods of economic recovery and downturns, such as the post-2008 crisis and the fluctuations around the COVID-19 pandemic. The seasonality component reveals repeating patterns within the data at regular intervals, typically related to seasonal effects.

In Figure 3, there is a clear, consistent cyclical pattern throughout the years, indicating that inflation rates follow a regular, periodic fluctuation. This pattern could be driven by factors such as seasonal changes in consumer demand, production cycles, or other recurring economic activities. The residual component represents irregular, random fluctuations in the data after removing the trend and seasonal effects. It captures noise or anomalies that are not explained by the trend or seasonality. The residual plot shows some significant spikes, particularly around the 2008 financial crisis and other economic shocks, suggesting periods of unexpected inflationary behavior. The relatively stable residuals outside of these spikes indicate that most of the variability in inflation can be explained by the identified trend and seasonal patterns. After training the various ML and traditional models, the results of the “in-sample” monthly predictions for the test data are depicted in Table 3.

Table 3. In-sample inflation predictions using machine learning and traditional models.

Month	Inflation	ARIMA	ETS	LinReg	XGBoost	AdaBoost	GB
2023-06-01	2.20	2.17	2.16	2.19	2.20	2.19	2.18
2023-07-01	2.30	2.15	2.15	2.26	2.23	2.18	2.18
2023-08-01	2.34	2.14	2.14	2.27	2.25	2.17	2.20
2023-09-01	2.34	2.12	2.09	2.28	2.30	2.18	2.24
2023-10-01	2.39	2.12	2.10	2.27	2.28	2.18	2.23
2023-11-01	2.30	2.11	2.13	2.29	2.25	2.18	2.22
2023-12-01	2.18	2.10	2.11	2.22	2.14	2.18	2.20
2024-01-01	2.27	2.10	2.14	2.26	2.16	2.21	2.23
2024-02-01	2.28	2.09	2.19	2.28	2.22	2.24	2.25
2024-03-01	2.31	2.09	2.23	2.28	2.18	2.26	2.25
2024-04-01	2.39	2.09	2.25	2.30	2.05	2.26	2.25
2024-05-01	2.33	2.08	2.21	2.33	2.04	2.25	2.22

4.2. Deployment of the transformer DL model

In this study, the transformer deep learning model was deployed by scaling the USA inflation data between 0 and 1 for optimal performance. The Min-Max scaling method was adopted for this operation. This setup ensures that the model receives normalized data, reducing potential biases. Next, we converted the scaled inflation time-series data into sequences that the transformer model can process. Here, a function that takes a sequence length as input and constructs overlapping subsequences from the data was defined. Each sequence consists of a specified number of consecutive data points (12 months of USA inflation used in this study) followed by the next point, which serves as the target value for the model to predict. These sequences were then transformed into PyTorch tensors as they are to be used in a transformer model, which was built using PyTorch. Extensive hyperparameter tuning resulting in the optimal hyperparameters (embedding dimension of 128, feed-forward hidden layer size of 64, dropout rate of 25%, number of epochs of 100, learning rate of 0.01, number of heads –5, number of layers –5, batch size of 128, etc.) were used to build the transformer model. The training was done using Adam optimizer and mean squared error as the loss function over the 100 epochs using DataLoader. Finally, the transformer model’s performance is evaluated by comparing predicted versus actual values and calculating error metrics like MAE, RMSE, MAPE, and SMAPE. Figure 5 presents the actual and predicted monthly inflation rates of the transformer deep-learning model.

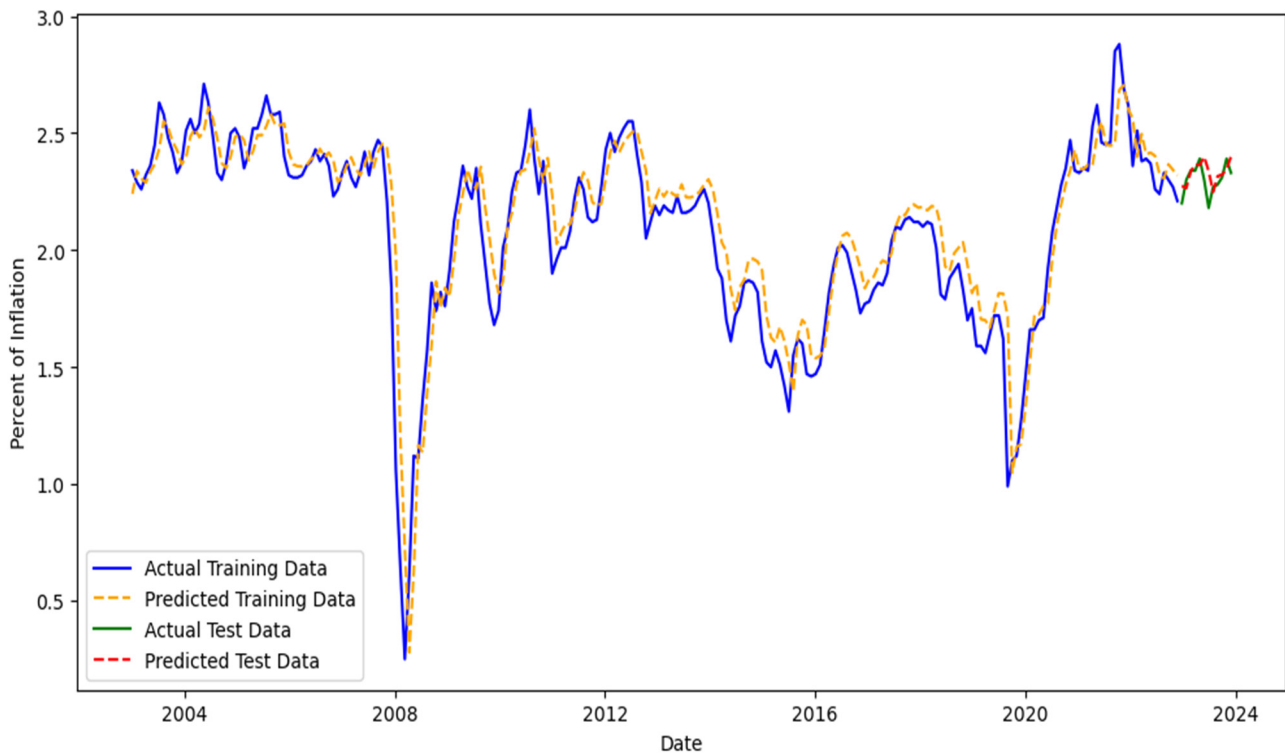


Figure 4. Actual and predicted monthly inflation rates by the transformer DL model.

The time-series plot presented in Figure 4 showcases the predictive performance of the transformer model over the training and testing dataset, presenting both the actual and predicted percentage of inflation over time. The actual training data demonstrates the observed values used during the model's training phase, while the predicted outcomes from the same phase closely align with the actual data, indicating an effective learning process of the transformer model. It is also observed that the predictions for the test data also follow the trend of the actual test data but display slight noticeable deviations, particularly at data points where sudden changes occur. This divergence highlights potential areas for improvement in the model's predictive accuracy and suggests that further tuning techniques might be necessary to enhance performance on unseen data.

4.3. Evaluation of ML, traditional statistical, and DL models

Table 4 presents the performance metrics for ML, traditional statistical, and DL models used to forecast monthly inflation. The metrics evaluated include MAE, RMSE, MAPE, and SMAPE. Among the models, the DL transformer model demonstrates the best overall performance with the lowest MAE (0.0221), RMSE (0.0291), MAPE (0.0288), and SMAPE (0.0281). This indicates that the DL transformer model provides the most accurate and consistent forecasts compared to the ML and traditional baseline models. Following the transformer model, linear regression also performs well, particularly with a low MAE (0.0437), RMSE (0.0554), MAPE (1.8753), and SMAPE (1.9010). The boosting models (GB, XGBoost, and AdaBoost) also show good performance, with moderate errors across all metrics considered in the study.

Table 4. Model performance metrics for traditional, ML, and DL models.

Model	MAE	RMSE	MAPE	SMAPE
ARIMA	0.1895	0.2038	8.1552	8.5555
ETS	0.1455	0.1619	6.2607	6.5115
Linear regression	0.0437	0.0554	1.8753	1.9010
GB	0.1116	0.1482	4.7839	4.9944
AdaBoost	0.0965	0.1152	4.1310	4.2555
XGBoost	0.0857	0.0984	3.6729	3.7626
Transformers	0.0221	0.0291	0.0288	0.0281

In contrast, both ARIMA and ETS models exhibited substantially higher error rates compared to the error rates from all the ML and DL models, indicating less reliable predictions of the USA inflation data. These findings suggest that all ML and DL models outperform the traditional baseline models (ARIMA and ETS) considered in this study. It is worth knowing that simpler linear ML models may outperform more complex ML techniques, as evident in this study (with the simple linear regression model outperforming all traditional and boosting algorithms), potentially due to the nature of the USA inflation data. According to the interpretation of typical MAPE values from Yıldırım et al. (2019), which was adopted for the study, it is evident that all the models (traditional, ML, and DL) had a high forecasting accuracy with all $\text{MAPE} \leq 10\%$. In lieu of this, all models were deemed fit to forecast the monthly inflation rate. The out-of-sample inflation forecasts for the ML and traditional models from June 2024 to December 2025 are given in Table 5.

Table 5. Out-of-sample inflation forecast using machine learning and traditional models.

Month	ARIMA	ETS	LinReg	XGBoost	AdaBoost	GB
2024-06-01	2.29	2.27	2.30	2.29	2.30	2.26
2024-07-01	2.25	2.27	2.29	2.23	2.25	2.16
2024-08-01	2.23	2.26	2.28	2.15	2.23	2.06
2024-09-01	2.20	2.21	2.24	2.11	2.22	1.99
2024-10-01	2.18	2.23	2.21	2.05	2.23	1.96
2024-11-01	2.17	2.25	2.20	2.01	2.23	1.96
2024-12-01	2.16	2.23	2.17	1.92	2.23	1.98
2025-01-01	2.14	2.26	2.21	1.92	2.28	2.02
2025-02-01	2.13	2.31	2.26	1.97	2.30	2.05
2025-03-01	2.13	2.35	2.26	2.01	2.33	2.04
2025-04-01	2.12	2.37	2.28	2.04	2.32	2.00
2025-05-01	2.11	2.33	2.29	1.99	2.31	1.92
2025-06-01	2.11	2.27	2.28	1.88	2.29	1.86
2025-07-01	2.11	2.27	2.29	1.83	2.27	1.82
2025-08-01	2.10	2.26	2.29	1.80	2.25	1.83
2025-09-01	2.10	2.21	2.25	1.76	2.24	1.86
2025-10-01	2.10	2.23	2.25	1.71	2.23	1.89
2025-11-01	2.09	2.25	2.23	1.73	2.22	1.89
2025-12-01	2.09	2.23	2.18	1.78	2.22	1.89

From Table 5 and Figure 5, it can be observed that the ARIMA model shows a slight initial rise in June 2024 (2.29%); then, it experiences minor fluctuations but there is a general downward trend after July 2024, stabilizing around 2.09% by December 2025. The ARIMA model predicts a steady

decrease in inflation over the period. The ETS forecast is relatively stable but shows a slight increase, peaking around March 2025 (2.35%). After this peak, the ETS model forecasts a gradual decline, reaching around 2.23% by December 2025. This model suggests moderate fluctuations with a general downward trend. Linear regression shows a stable forecast with a minor initial rise to around August 2024 (2.28%). The model predicts a slight decrease, with the value dropping to 2.18% by December 2025. The overall trend is relatively flat, indicating minimal changes over time. The XGBoost model forecasts a noticeable rise in July 2024 (2.23%), followed by minor increases. A sharp decline is predicted for October 2024 (2.05%), with further reductions reaching 1.78% by December 2025. The XGBoost model shows significant short-term fluctuations.

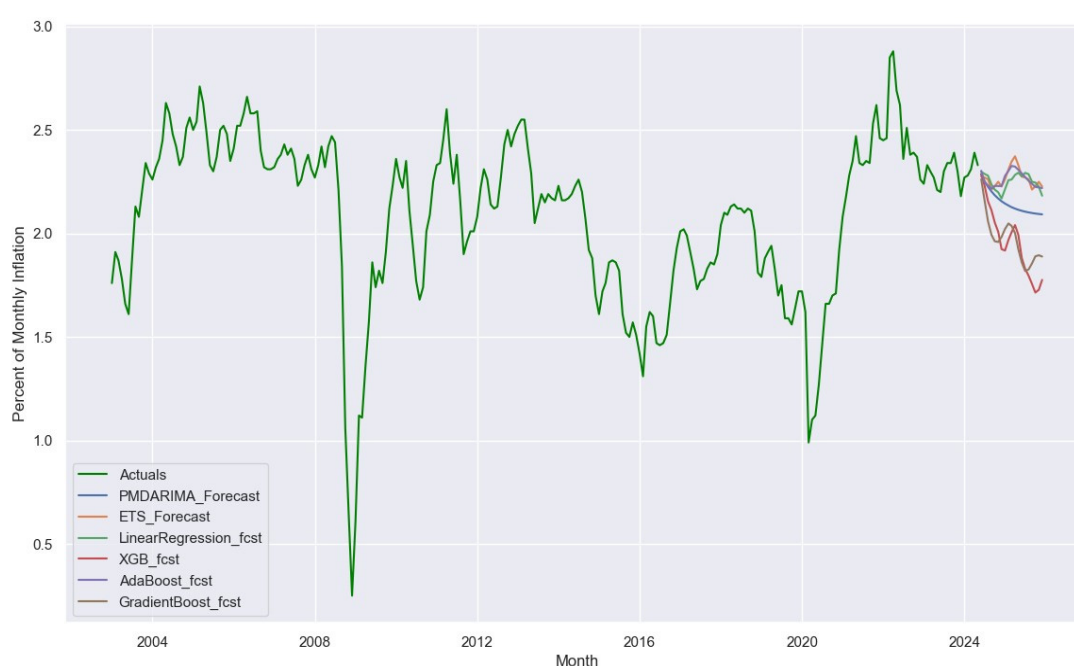


Figure 5. Monthly forecast inflation rates of both machine learning and traditional models.

AdaBoost shows an initial increase, peaking in June 2024 (2.30%), and then fluctuates slightly. There is a marked decrease starting in July 2024, stabilizing around 2.22% by December 2025. The AdaBoost model forecasts a declining trend with minor ups and downs. The gradient boosting model has an initial rise in June 2024 (2.26%), followed by minor increases and decreases. The GB model shows a more consistent drop starting from August 2024, reaching 1.89% by December 2025, and predicts a steady decline over the forecast period. In terms of consistency, linear regression and ETS models show the most stable forecasts with minimal fluctuations, suggesting they might be more reliable for steady, long-term forecasting. Regarding volatility, XGBoost and AdaBoost exhibit more volatility, indicating sensitivity to short-term changes and possibly overfitting to recent trends. In general, almost all traditional and ML models considered in this study predict a decline in inflation over the forecast period (June 2024 to December 2025), with varying degrees of confidence and fluctuation.

4.4. Discussion of results

The two traditional statistical models that have long been regarded as the backbone of time-series forecasting are ARIMA and ETS. The strong theoretical foundations of these two models make them relevant to date. In this study, however, both ARIMA and ETS produced high error metrics in comparison to the ML and DL models considered. Specifically, high MAE, RMSE, MAPE, and SMAPE values were observed for ARIMA and ETS, which is indicative of less reliable inflation forecasts for the USA data. With the complexity and nonlinearity inherent to the USA inflation data, the well-known ARIMA model, recognized for its effectiveness in capturing linear relationships and short-term dependencies, struggled to produce better forecasts. Likewise, the ETS model, despite its ability to handle seasonality and trends, also fell short in providing robust inflation forecasts for the USA data. Aside from this study, ML and DL models adopted in other settings have also demonstrated superior performance over conventional baseline traditional statistical models. For instance, a study on Brazil's inflation forecasting Araujo and Gaglianone (2023) revealed that ML methods, such as random forest and GB were superior to traditional econometric models. Similarly, a study on USA inflation forecasting Paranhos (2025), which compared DL models with traditional models, found that DL models provided better forecasting and were better at capturing temporal dependencies and nonlinear patterns. It is then prudent that for effective economic planning and decision-making, researchers can rely on the superior performance of ML and DL models in their estimations.

There are other notable studies in the literature, such as Aygun and Gunay (2021) who adopted traditional techniques and ML algorithms for forecasting daily Bitcoin returns and recommended ML algorithms due to their robust performance compared to traditional models. Another study Medeiros et al. (2021) focused on forecasting inflation in a data-rich environment using ML models and presented significant advancements in the field. By comparing traditional linear models with modern nonlinear ML models, the study demonstrated the superiority of nonlinear ML models, such as random forests, for inflation forecasting over traditional benchmark models. Furthermore, a study compared predictive models for forecasting Ecuador's Consumer Price Index (CPI), Riofrío et al. (2020) found that ML and DL algorithms, specifically SVR with a polynomial kernel and LSTM neural networks, outperformed traditional models such as SARIMA and ETS. The SVR, with a polynomial kernel, achieved the lowest MAPE of 0.00171, followed closely by the LSTM neural network with a MAPE of 0.00173.

Likewise, another study investigated whether ML models can outperform traditional models in forecasting USA's gross domestic product (GDP). By comparing classical time-series models (AR and SARIMA) and ML algorithms (KNN and linear regression), the findings of Maccarrone et al. (2021) suggest that the ML KNN model outperforms traditional models in one-step-ahead forecasts, leveraging repetitive patterns within the time-series data. However, for longer forecasting horizons, the performance of KNN diminishes, and models incorporating financial variables, such as linear regression with the yield curve, show superior accuracy. These findings from other researchers in the literature suggest that advanced ML and DL models can provide more accurate forecasts for economics and financial time-series data than conventional traditional approaches.

The comparison between this study's findings with other relevant similar studies in the literature that used economics datasets is presented in Table 6.

Table 6. Comparison of RMSE and MAE of our study with other studies in the literature that used economics datasets.

Research work	Techniques adopted	Dataset used	Best error metrics
Dong (2020)	Random walk (RW), two-pillar Philips curve (multi-variate model), ARIMA, VAR, and VECM	China inflation	RMSE (0.754) and MAE (0.570) with VECM
Ingabire and Mung'atu (2016)	ARIMA (3,1,4) and VECM (2,2)	Rwanda inflation	RMSE (0.980) and MAE (1.580) with VECM (2,2)
Wang et al. (2024)	ARIMA, random forest, SVM regression, and SVR	USA healthcare expenditure	RMSE (0.297) with random forest
Nortey et al. (2024)	ARMA, GARCH, ARMA-GARCH, and GARCH-MIDAS	Ghana, Nigeria, and South Africa stock	RMSE (4.423) and MAE (0.970) with GARCH-MIDAS
Jouilil and Iaousse (2023)	ARIMA-GARCH, ETS, KNN, Prophet, and LSTM	USA inflation	RMSE (0.756) and MAE (0.629) with KNN
Jha et al. (2024)	Ridge, LASSO, SVR with both linear and RBF kernel, MR with forward, backward, and best subset selection	WTI crude oil spot prices	RMSE (0.0766) with SVR with RBF kernel
Zhao et al. (2017)	RW, MRS, SVR, SVR-B, FNN, FNN-B, SDAE, and SDAE-B	WTI crude oil spot prices	RMSE (4.995) with SDAE-B
Huang and Wang (2018)	WNN, WNNRT, and SVM	WTI and brent crude oil spot prices	RMSE (2.050) and MAE (0.970) with WNNRT
Cen and Wang (2019)	LSTM, NLP, RNN, and EEMD	WTI and brent crude oil spot prices	0.450 (RMSE) and 0.18 (MAE) with LSTM
Busari and Lim (2021)	LSTM, GRU, AdaBoost-LSTM, and AdaBoost-GRU	Export crude oil price	RMSE (2.460) and MAE (1.416) with AdaBoost-GRU
Aras and Lisboa (2022)	RF, LASSO, GB, XGBoost, AdaBoost, ERT	TURKSTAT, EUROSTAT, CBRT, and IMF	RMSE (0.815) with AdaBoost and 0.931 (MAE) with GB
Joseph et al. (2024)	RF, LASSO, Boosted Trees, DNN, AdaLASSO	UK inflation	0.710 (RMSE) and 0.670 (MAE) with RF
Alomani et al. (2025)	ARIMA and GB	Global inflation	0.072 (RMSE) and 0.300 (MAE) with GB
Özgür and Akkoç (2022)	ARIMA (1,1,0), ARIMA (3,1,2), MVAR, FB Prophet, Ridge, Lasso, AdaLASSO, Group Lasso, and Elastic Net	Turkey inflation	0.084 (RMSE) with lasso
Our Method	ETS, ARIMA, Linear Regression, GB, XGBoost, AdaBoost, and Transformer	USA inflation	0.029 (RMSE) and 0.022 (MAE) with transformers

NB: VECM: vector error correction model; MR: multivariate regression; WTI: West Texas intermediate; SAE: stacked autoencoders; DAEs: denoising autoencoders; WNNRT: Wavelet neural network with random time effective function; EEMD: ensemble empirical mode decomposition; MVAR: multivariate vector autoregression; NLP: Natural language processing; FNN: Feedforward Neural Network; FB: Facebook; ERT: Extremely randomized trees.

5. Conclusions and recommendations

This study evaluated the performance of baseline traditional statistical, ML, and DL models for predicting monthly inflation rates in the USA. Overall, the results of this study showed consistent performance improvements of DL and ML-based models like transformers, linear regression, GB, XGBoost, and AdaBoost when compared with traditional benchmark models like ARIMA and ETS. The debate over whether DL can outperform both ML and traditional models for time-series forecasting is ongoing, but it should be noted that all approaches do have their different pros and cons.

The performance of any model will depend on a variety of factors, such as the characteristics of the data (seasonality, trend, residuals, etc.), the nature of the forecasting problem, model hyperparameters, available computational resources, interpretability requirements, and the trade-offs between simplicity and performance. In the context of time-series data, traditional models are known for their ability to capture linear relationships, seasonal behavior, and autocorrelations. Such models are effective due to their simple implementation and are well-understood theoretically, thus making them reliable for a host of prediction tasks.

In contrast, as complex nonlinear relationships and patterns can be captured using ML and DL models, they became popular. Indeed, these models, including transformers, linear regression, GB, XGBoost, and AdaBoost, have shown competitive performance against traditional models, as evidenced by this study. DL and ML algorithms have been shown to yield better forecasts than ARIMA and ETS models across popular benchmark datasets (Wang et al., 2024). Ingabire and Mung'atu (2016) also indicated that, in many cases, ML and DL models outperform traditional baseline models. Nevertheless, incorporating DL and ML models into a forecasting toolkit is a non-trivial task and requires careful consideration of features, as this process is paramount to model performance. These carefully selected features can be fine-tuned and supplemented with a thorough model hyperparameter optimization process to further support the accuracy and reliability of forecasts. As such, given effective fine-tuning, DL and ML models can be powerful contributors to the forecasting toolkit.

However, it is worth noting that for a variety of time series data, traditional concepts remain useful, and both ML and DL improvements are particularly suited for the identification of complex nonlinear variable interactions, thus enabling more accurate forecasting. The choice between benchmark traditional statistical, ML, and DL models is solely dependent on the forecasting task and the nature of the data at hand. By utilizing the strengths of baseline traditional statistical, ML, and DL models, researchers can achieve more reliable and accurate predictions. Future research should explore hybrid models to improve economics forecasting accuracy.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author contributions

The authors declare to have contributed equally to the manuscript. All authors read and approved of the final manuscript.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The second author acknowledges the enormous support of the University of Texas Rio Grande Valley (UTRGV) Presidential Research Fellowship fund.

Conflict of interest

The authors declare no conflict of interest.

References

- Adnan RM, Liang Z, Kuriqi A, et al. (2021) Air temperature prediction using different machine learning models. *Indones J Electr Eng Comput Sci* 22: 534–541. <https://doi.org/10.11591/ijeecs.v22.i1.pp534-541>
- Agyemang EF, Mensah JA, Ocran E, et al. (2023) Time series based road traffic accidents forecasting via SARIMA and Facebook Prophet model with potential changepoints. *Heliyon* 9: 1–18. <https://doi.org/10.1016/j.heliyon.2023.e22544>
- Alomani G, Kayid M, Abd El-Aal MF (2025) Global inflation forecasting and Uncertainty Assessment: Comparing ARIMA with advanced machine learning. *J Radiat Res Appl Sci* 18: 1–7. <https://doi.org/10.1016/j.jrras.2025.101402>
- Aras S, Lisboa PJ (2022) Explainable inflation forecasts by machine learning models. *Expert Syst Appl* 207: 1–19. <https://doi.org/10.1016/j.eswa.2022.117982>
- Araujo GS, Gaglianone WP (2023) Machine learning methods for inflation forecasting in Brazil: New contenders versus classical models. *Lat Am J Cent Bank* 4: 1–29. <https://doi.org/10.1016/j.latcb.2023.100087>
- Arora S, Rani R, Saxena N (2024) A systematic review on detection and adaptation of concept drift in streaming data using machine learning techniques. *Wires Data Min Knowl* 14: 1–27. <https://doi.org/10.1002/widm.1536>
- Aygun B, Gunay EK (2021) Comparison of statistical and machine learning algorithms for forecasting daily bitcoin returns. *Avrupa Bilim ve Teknoloji Dergisi* 21: 444–454. <https://doi.org/10.31590/ejosat.822153>
- Ben Bouallègue Z, Clare MC, Magnusson L, et al. (2024) The rise of data-driven weather forecasting: A first statistical assessment of machine learning–based weather forecasts in an operational-like context. *B Am Meteorol Soc* 105: 864–883. <https://doi.org/10.1175/BAMS-D-23-0162.1>
- Bonci A, Kermenov R, Longarini L, et al. (2024) Artificial Intelligence in Manufacturing: A Lightweight Framework for Online Anomaly Detection at the Edge. *Commun Ind* 2024: 307–311.
- Busari GA, Lim DH (2021) Crude oil price prediction: A comparison between AdaBoost-LSTM and AdaBoost-GRU for improving forecasting performance. *Comput Chem Eng* 155: 1–9. <https://doi.org/10.1016/j.compchemeng.2021.107513>
- Cen Z, Wang J (2019) Crude oil price prediction model with long short term memory deep learning based on prior knowledge data transfer. *Energy* 169: 160–171. <https://doi.org/10.1016/j.energy.2018.12.016>
- Dong J (2020) Forecasting modeling for China's inflation. Working paper.
- Gunnarsson ES, Isern HR, Kaloudis A, et al. (2024) Prediction of realized volatility and implied volatility indices using AI and machine learning: A review. *Int Rev Financ Anal* 103221: 1–20. <https://doi.org/10.1016/j.irfa.2024.103221>
- He QQ, Pang PCI, Si YW (2019) Transfer learning for financial time series forecasting. In *PRICAI 2019: Trends in Artificial Intelligence: 16th Pacific Rim International Conference on Artificial*

- Intelligence*, Springer International Publishing 1: 24–36. https://doi.org/10.1007/978-3-030-29911-8_3
- Huang L, Wang J (2018) Global crude oil price prediction and synchronization based accuracy evaluation using random wavelet neural network. *Energy* 151: 875–888. <https://doi.org/10.1016/j.energy.2018.03.099>
- Ingabire J, Mung'atu JK (2016) Measuring the performance of autoregressive integrated moving average and vector autoregressive models in forecasting inflation rate in Rwanda. *Int J Math Phys Sci Res* 4: 15–25.
- Jha N, Tanneru HK, Palla S, et al. (2024) Multivariate analysis and forecasting of the crude oil prices: Part I–Classical machine learning approaches. *Energy* 296: 1–13. <https://doi.org/10.1016/j.energy.2024.131185>
- Joseph A, Potjagailo G, Chakraborty C, et al. (2024) Forecasting UK inflation bottom up. *Int J Forecast* 40: 1521–1538. <https://doi.org/10.1016/j.ijforecast.2024.01.001>
- Jouilil Y, Iaousse MB (2023) Comparing the accuracy of classical and machine learning methods in time series forecasting: A case study of usa inflation. *Stat Optim Inform Comput* 11: 1041–1050. <https://doi.org/10.19139/soic-2310-5070-1767>
- Kontopoulou VI, Panagopoulos AD, Kakkos I, et al. (2023) A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. *Future Internet* 15: 2–31. <https://doi.org/10.3390/fi15080255>
- Maccarrone G, Morelli G, Spadaccini S (2021) GDP forecasting: Machine learning, linear or autoregression? *Front Artif Intell* 4: 1–19. <https://doi.org/10.3389/frai.2021.757864>
- Makala D, Li Z (2021) Prediction of gold price with ARIMA and SVM. *Journal of Physics: Conference Series* 1767: 1–9. <https://doi.org/10.1088/1742-6596/1767/1/012022>
- Masini RP, Medeiros MC, Mendes EF (2023) Machine learning advances for time series forecasting. *J Econ Surv* 37: 76–111. <https://doi.org/10.1111/joes.12429>
- Medeiros MC, Vasconcelos GF, Veiga Á, et al. (2021) Forecasting inflation in a data-rich environment: the benefits of machine learning methods. *J Bus Econ Stat* 39: 98–119. <https://doi.org/10.1080/07350015.2019.1637745>
- Nortey EN, Agbeli R, Debrah G, et al. (2024) A GARCH-MIDAS approach to modelling stock returns. *Commun Stat Appl Met* 31: 535–556. <https://doi.org/10.29220/CSAM.2024.31.5.535>
- Özgür Ö, Akkoç U (2022) Inflation forecasting in an emerging economy: selecting variables with machine learning algorithms. *Int J Emerg Mark* 17: 1889–1908. <https://doi.org/10.1108/IJOEM-05-2020-0577>
- Paranhos L (2025) Predicting inflation with recurrent neural networks. *Int J Forecast* 5283: 1–16. <https://doi.org/10.1016/j.ijforecast.2024.07.010>
- Reed SB (2014) One hundred years of price change: The Consumer Price Index and the American inflation experience. *Monthly Lab Rev* 137: 1–32. <https://doi.org/10.21916/mlr.2014.14>
- Riofrío J, Chang O, Revelo-Fuelagán EJ, et al. (2020) Forecasting the Consumer Price Index (CPI) of Ecuador: A comparative study of predictive models. *Int J Adv Sci Eng Inf Technol* 10: 1078–1084. <https://doi.org/10.18517/ijaseit.10.3.10813>
- Sezer OB, Gudelek MU, Ozbayoglu AM (2020) Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Appl Soft Comput* 90: 1–32. <https://doi.org/10.1016/j.asoc.2020.106181>

- Valderrama Balaguera JC (2024) Precipitation forecast estimation applying the change point method and ARIMA. *Cogent Eng* 11: 1–13. <https://doi.org/10.1080/23311916.2024.2340191>
- Vaughan L, Zhang M, Gu H, et al. (2023) An exploration of challenges associated with machine learning for time series forecasting of COVID-19 community spread using wastewater-based epidemiological data. *Sci Total Environ* 858: 1–8. <https://doi.org/10.1016/j.scitotenv.2022.159748>
- Wang J, Qin Z, Hsu J, et al. (2024) A fusion of machine learning algorithms and traditional statistical forecasting models for analyzing American healthcare expenditure. *Healthcare Anal* 5: 1–10. <https://doi.org/10.1016/j.health.2024.100312>
- Westergaard G, Erden U, Mateo OA, et al. (2024) Time series forecasting utilizing automated machine learning (AutoML): A comparative analysis study on diverse datasets. *Information* 15: 1–20. <https://doi.org/10.3390/info15010039>
- Yıldırım S, Tosun E, Çalık A, et al. (2019) Artificial intelligence techniques for the vibration, noise, and emission characteristics of a hydrogen-enriched diesel engine. *Energ Source Part A* 41: 2194–2206. <https://doi.org/10.1080/15567036.2018.1550540>
- Zhang X, Qi W, Zhan Z (2021) A Study on Machine-Learning-Based Prediction for Bitcoin's Price via Using LSTM and SVR. *Journal of Physics: Conference Series* 1732: 1–5. <https://doi.org/10.1088/1742-6596/1732/1/012027>
- Zhao Y, Li J, Yu L (2017) A deep learning ensemble approach for crude oil price forecasting. *Energy Econ* 66: 9–16. <https://doi.org/10.1016/j.eneco.2017.05.023>



AIMS Press

© 2025 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0>)