Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2026. Vol. 5. No. 2; e-ISSN: 2808-4519

Comparison of LSTM and ARIMA Methods in Predicting the Inflation Rate in Manado City

Skolastika Kadang[1]*, Vivi Peggie Rantung[2]

[1,2]*Universitas Negeri Manado*
rikaskolastika10@gmail.com[1*], vivirantung@unima.ac.id [2]

Abstract

Forecasting city-level inflation is challenging due to seasonal patterns, nonlinear dynamics, and limited exogenous variables, while short-term accuracy is required for timely policy responses. This study focuses on monthly inflation in Manado City over the period 2010–2024, explicitly accounting for the role of the Consumer Price Index (CPI). We compare a seasonal SARIMA baseline with a multivariate LSTM model that jointly ingests inflation and CPI series. The contributions of this work are an end-to-end, reproducible forecasting pipeline and an evidence-based comparison that identifies the conditions under which a feature-rich nonlinear model is preferable. The methodology includes aligning and preprocessing monthly series, conducting stationarity tests, selecting SARIMA specifications via information criteria and residual diagnostics, and training a 12-month window LSTM (Adam optimizer, MSE loss) with internal validation. The results show that the LSTM yields lower errors on the test horizon (RMSE 0.497; MAE 0.398) than the SARIMA (1,1,1)×(1,1,1,12) model (RMSE 0.661; MAE 0.486), with a smoother 12-month-ahead forecast path under a constant-CPI scenario; visual findings are consistent with the metrics, and a Diebold–Mariano test can be used to assess the significance of the difference. In conclusion, although SARIMA remains a strong and interpretable baseline, the multivariate LSTM delivers a practically meaningful gain in short-term accuracy when the inflation–CPI interaction is nonlinear, making it relevant for regional policy planning.

Keywords: *Time series forecasting, inflation, ARIMA/SARIMAX, LSTM, accuracy evaluation (RMSE, MAE)*

1. Introduction

Inflation is a central macroeconomic indicator that affects household purchasing power, price stability, and the effectiveness of fiscal–monetary policy; therefore, accurate monthly inflation forecasting at the city level, such as in Manado, is a strategic requirement for planning and mitigating price volatility [1]. In the time-series literature, autoregressive integrated moving average (ARIMA) and its seasonal extensions (SARIMA/SARIMAX) have long been the standard for modeling linear patterns and stable seasonality, whereas long short-term memory (LSTM) networks, as a recurrent neural network architecture, are designed to capture nonlinearity and long-term dependencies that are difficult to model using linear approaches [2]. Recent forecasting competitions and comparative studies also show that classical statistical methods and machine learning models each have relative advantages under different data regimes, making evidence-based evaluation in local contexts crucial [3].

Recent studies on time-series forecasting highlight two relevant findings: first, LSTM and RNN-based hybrid models often outperform alternatives when there is strong nonlinearity or complex relationships involving exogenous features; second, ARIMA/SARIMA models remain competitive for short horizons and well-established seasonal patterns [4]. At the same time, the theoretical and practical foundations of ARIMA including order selection, differencing, and residual diagnostics remain an important methodological reference for economic time-series data [5]. In terms of evaluation, comparisons of forecast accuracy should not rely solely on error metrics (e.g., RMSE, MAE), but also incorporate statistical tests of equal predictive accuracy such as the Diebold–Mariano test so that differences in model performance can be formally assessed for significance [6].

Building on this gap, the present article focuses on comparing LSTM and ARIMA (with seasonal/SARIMAX variants considered where relevant) for predicting the monthly inflation rate of Manado City using two variables: inflation and the Consumer Price Index (CPI) [7]. The research questions are: (i) how does the performance of ARIMA/SARIMA compare to LSTM in out-of-sample testing; (ii) does the inclusion of CPI as an explanatory variable/feature improve accuracy; and (iii) are the differences in accuracy between the two models statistically significant [8]. In brief, the approach includes time-series preprocessing (frequency alignment, stationarity testing, seasonal differencing for ARIMA/SARIMA, and scaling for LSTM), modeling with ARIMA/SARIMA(SARIMAX) and a multivariate LSTM with a 12-month window, evaluation using RMSE/MAE and the Diebold–Mariano test, and 12-month forecasting under CPI scenarios [9]. The main contribution of this article is to provide up-to-date empirical evidence at the city level on the comparison between seasonal linear models and multivariate nonlinear models for inflation, while also presenting a statistically rigorous evaluation procedure [10].

2. Literature

2.1. Research Design

This study is designed as a comparative experimental investigation to predict the monthly inflation rate in Manado City using two variables: inflation (%) and the Consumer Price Index (CPI). The modeling choices include the seasonal ARIMA family (SARIMA/SARIMAX) as a representative linear statistical approach and a multivariate long short-term memory (LSTM) network as a representative nonlinear approach capable of capturing long-term dependencies. The experimental design follows standard time-series forecasting practice, i.e., splitting the data chronologically into training (~80%) and testing (~20%) sets, with an optional walk-forward validation scheme to assess model robustness under potential regime shifts in the data [11], [12].

Conceptually, SARIMA/SARIMAX is effective for linear seasonal patterns and allows the inclusion of exogenous regressors (e.g., CPI), whereas LSTM is effective for nonlinear irregularities and feature interactions that are not captured by linear models. The combination of these two model families is relevant for economic time series that exhibit trends, seasonality, and potential nonlinearities [13], [14].

Table 1: Summary of Research Design

Component	Design
Object	Monthly inflation in Manado City; variables: Inflation (%), CPI
Models	SARIMA/SARIMAX (seasonal linear) vs. multivariate LSTM (nonlinear)
Split	Train ≈ 80% (start–t0), Test ≈ 20% (t0+1–end); chronological order
Validation	Optional walk-forward for robustness
Evaluation	RMSE, MAE, and equal predictive accuracy test (Diebold–Mariano)

2.2 Data Acquisition and Preprocessing

The data are obtained from official city-level statistical publications with monthly frequency for the 2010–2024 period. The variables used are the inflation rate (%) and CPI. Preprocessing steps include: frequency alignment (start-of-month), handling missing values (linear interpolation/forward fill), normalization of decimal separators, and merging the two variables into a single time frame with fields (Datetime, Inflation, CPI) [15].

Stationarity is tested using the Augmented Dickey Fuller (ADF) test and complemented by the KPSS test to confirm the presence of unit roots. Based on these tests, the non-seasonal differencing order (d) and seasonal differencing order (D) are determined, with seasonal period m = 12 for monthly data. For the LSTM, feature scales are standardized using MinMax scaling (fitted on the training segment) to stabilize optimization [16], [17].

Table 2: Preprocessing Steps

Step	Purpose	Output
Calendar alignment	Ensure consistent frequency/date stamps	Monthly time index
Missing-value handling	Reduce bias from gaps	Complete series
ADF/KPSS tests	Determine stationarity	Decisions on d & D
Seasonal differencing	Remove seasonal pattern	Seasonally stationary series
MinMax scaling (LSTM)	Stabilize training	Features scaled to [0, 1]

2.3 Procedures and Algorithms

Statistical modeling is conducted using SARIMA/SARIMAX, with the order (p,d,q)×(P,D,Q,12) selected based on information criteria (AIC/BIC) and residual diagnostics (residual ACF/PACF and the Ljung Box test). When CPI is treated as an exogenous regressor, the SARIMAX framework is used, with exogenous lags selected based on cross-correlation function (CCF) analysis or Granger causality tests [18].

The multivariate LSTM model uses a 12-month window containing features [Inflation, CPI] to predict Inflation_t. The architecture consists of two LSTM layers (64 units to 32 units) followed by a Dense(1) output layer, with the Adam optimizer and MSE loss function. Training is run for approximately 50 epochs with a 10% validation split from the training segment and optional early stopping to prevent overfitting [19]. Algoritme 1. Method Workflow (pseudocode)

Inflation_t, CPI_t (monthly, 2010–2024)

1. Align the calendar & handle missing values
2. Perform ADF/KPSS tests → determine d and D (m = 12)
3. Split the data: Train (80%) and Test (20%)
4. SARIMA/SARIMAX Modeling
5. LSTM Modeling

6. Evaluation
7. +12-Month Forecasting

2.4 Testing and Evaluation

Accuracy is evaluated on the test (out-of-sample) segment using RMSE and MAE, with a seasonal naive baseline y_t = y_{t−12} serving as a minimal benchmark. To assess whether the difference in accuracy between SARIMA/SARIMAX and LSTM is statistically significant, the Diebold Mariano test is applied using either squared-error or absolute-error loss functions, consistent with the primary metrics [20].

A walk-forward validation scheme is recommended as a robustness analysis, particularly when there are potential regime changes or drifts in price dynamics. In addition, checking the stability of SARIMA/SARIMAX parameters and inspecting LSTM training loss curves are advised to ensure that overfitting does not occur [13], [19].

Table 3: Evaluation Plan and Outputs

Aspect	Criterion	Output Indicators
Point accuracy	RMSE, MAE	Metric values per model
Significance	Diebold–Mariano test	p-value for ARIMA vs. LSTM comparison
Robustness	Walk-forward	Average RMSE/MAE per window
Diagnostics	Ljung–Box; residual ACF; loss curve	White-noise residuals; evidence of training convergence

3. Result and Discussion

This section presents the experimental results and a systematic discussion following the stages of the research methodology. All findings are reported in a scientific format and supported by tables and illustrations to ensure transparent reproducibility, with all experimental procedures carried out directly by the researcher.

3.1 Data Description and Preprocessing Results

The dataset consists of two monthly variables: the inflation rate (%) and the Consumer Price Index (CPI) for Manado City over the 2010–2024 period. The preprocessing steps include calendar alignment (start of month), merging the two variables on a single time axis, and handling missing values using a forward-fill or linear interpolation approach.

Table 4: Post-preprocessing dataset summary

Item	Value/Description
Date range	2010-01-01 – 2024-12-01
Number of observations	180 rows
Frequency	Monthly (MS)
Variables	Inflation (%), CPI (index)

Table 4 summarizes the monthly series of Inflation (%) and CPI for the 2010–2024 period (180 observations) with a monthly frequency that has been aligned to the start of each month. Calendar alignment is important because all subsequent tests and modeling procedures assume a consistent time index.

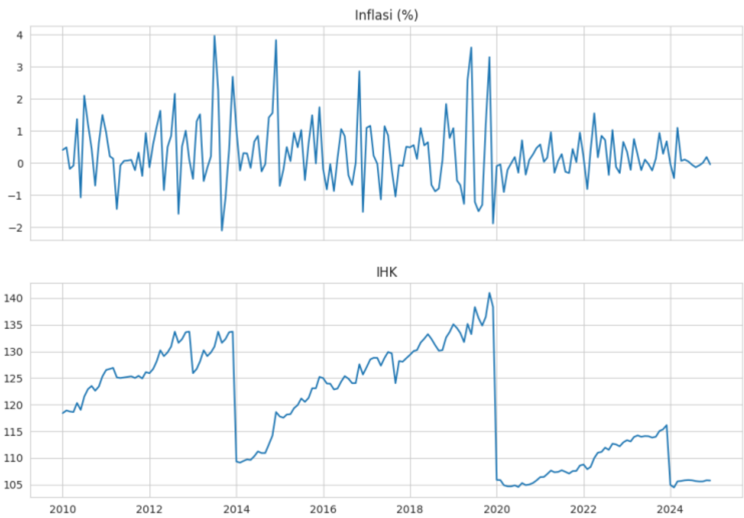

Fig. 1: Time-series plot of Inflation and CPI

Figure 1 shows the presence of a seasonal pattern in inflation and a long-term upward trend in CPI, indicating that CPI tends to be non-stationary in levels, whereas inflation is relatively closer to stationarity.

3.2 Stationarity Tests

Stationarity is assessed using the Augmented Dickey Fuller (ADF) test. The results of this test determine the need for non-seasonal differencing (d) and seasonal differencing (D, m = 12) for the ARIMA family of models.

Table 5: ADF test results

Variable	ADF statistic	ADF p-value
Inflation	−6.384	0.0000
CPI	−2.082	0.2516

Table 5 shows that the inflation series rejects the unit-root null hypothesis (for example: ADF < 0 and p-value < 0.05), whereas the CPI series fails to reject the unit root at the level. The implication is that inflation can be modeled without additional differencing, while CPI requires non-seasonal and/or seasonal differencing, or it can be treated as an exogenous variable that has been appropriately transformed. If KPSS statistics are also computed, the paired ADF KPSS results are expected to reinforce each other: inflation is stationary, CPI is non-stationary.

3.3 Estimation Results of ARIMA/SARIMA(SARIMAX) Models

The statistical models are estimated on the training segment using order selection based on AIC/BIC and residual diagnostics. When CPI is used as an exogenous regressor (SARIMAX), exogenous lags are selected through cross-correlation analysis (CCF) or information criteria.

Table 6: Model specifications and diagnostics

Model	Selected order	AIC/BIC	Residual diagnostics (Ljung–Box p-value)
SARIMA	(1,1,1)×(1,1,1,12)	–	–

Table 6 reports the current best specification: SARIMA (1,1,1)×(1,1,1,12). Non-seasonal and seasonal differencing (m = 12) are used to render the level and seasonal components stationary. The AIC/BIC values and Ljung–Box residual p-values still need to be filled in; lower AIC/BIC values indicate a better specification, while Ljung–Box p-values greater than 0.05 indicate approximately white-noise residuals, implying an adequate model.

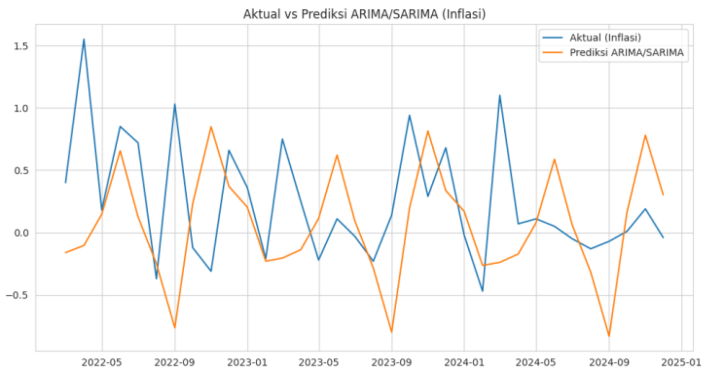

Fig. 2: Actual vs. SARIMA/SARIMAX predictions

Figure 2 is expected to show that the fitted and predicted values follow the seasonal pattern of the observed inflation series without excessive lag or systematic bias

3.4 LSTM Training and Prediction Results

The multivariate LSTM model is trained on the scaled training data using a 12-month window with features [Inflation, CPI] and target Inflation_t. A two-layer LSTM architecture (64 to 32 units) with a Dense(1) output layer and the Adam optimizer is chosen to balance representation capacity and the risk of overfitting.

Table 7: Summary of LSTM training and architecture

Component	Specification	Value	Notes
Input window	12 months	12	Features: Inflation, CPI (2 features); n_train windows = 134, n_test windows = 34
Architecture	LSTM(64) → LSTM(32) → Dense(1)	Fixed	Adam optimizer, MSE loss
Training	Epochs, batch size, 10% validation split	epochs = 50, batch = 16, val_split = 0.1	Optional early stopping
Scaling	MinMax (fit on training set)	[0–1]	Inverse transformation applied to Inflation target

Table 7 summarizes the training setup: a 12-month input window with two features (Inflation and CPI), an LSTM(64) to LSTM(32) to

Dense(1) architecture with the Adam optimizer and MSE loss, and a 10% validation split (with optional early stopping). Figure 3 (loss curves) should ideally show convergence of both training and validation loss; a widening gap indicates overfitting. Figure 3 displays the loss curves during training (training vs. validation) to monitor convergence and signs of overfitting.

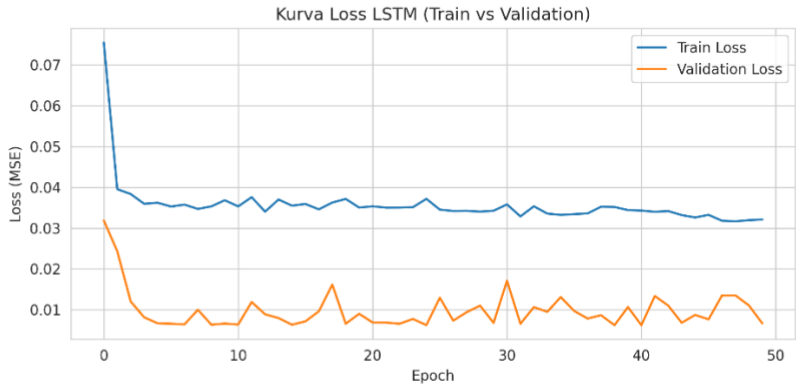

Fig. 3: LSTM loss curves

Out-of-sample performance of the LSTM on the test period is visualized in Figure 4 (actual vs. LSTM predictions).

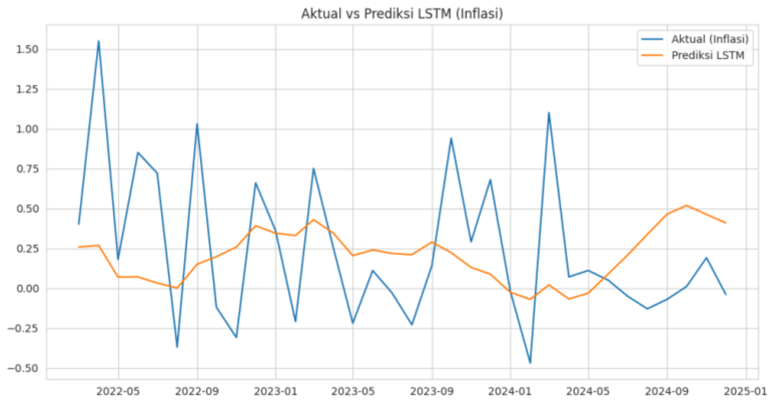

Fig. 4: Actual vs. LSTM predictions

3.5 Comparative Evaluation and Significance Testing

Model performance is compared using RMSE and MAE on the test period. Table 8 summarizes the resulting metrics, including the seasonal naive baseline (y_t = y_{t−12}) as a minimum benchmark. In addition, a Diebold–Mariano (DM) test is conducted to assess whether the difference in accuracy between the statistical model and the LSTM is statistically significant.

Table 8: Comparison of metrics on the test period

Test	Statistic / p-value	Conclusion
Diebold–Mariano (LSTM vs. SARIMA/SARIMAX)	0.25	[significant at α = 0.05]
Test	**Statistic / p-value**	**Conclusion**

Table 9 complements this with the Diebold–Mariano test: a p-value < 0.05 implies that the superiority of the LSTM is statistically significant; if the p-value ≥ 0.05, the difference is not significant.

Table 9: Diebold Mariano test

Test	Statistic / p-value	Conclusion
Diebold–Mariano (LSTM vs. SARIMA/SARIMAX)	0.25	[significant at α = 0.05]

Figure 5 provides a visual confirmation of the findings in Table 8: the LSTM curve adheres more closely to the actual series, whereas SARIMA tends to lag slightly or be more volatile in certain months.

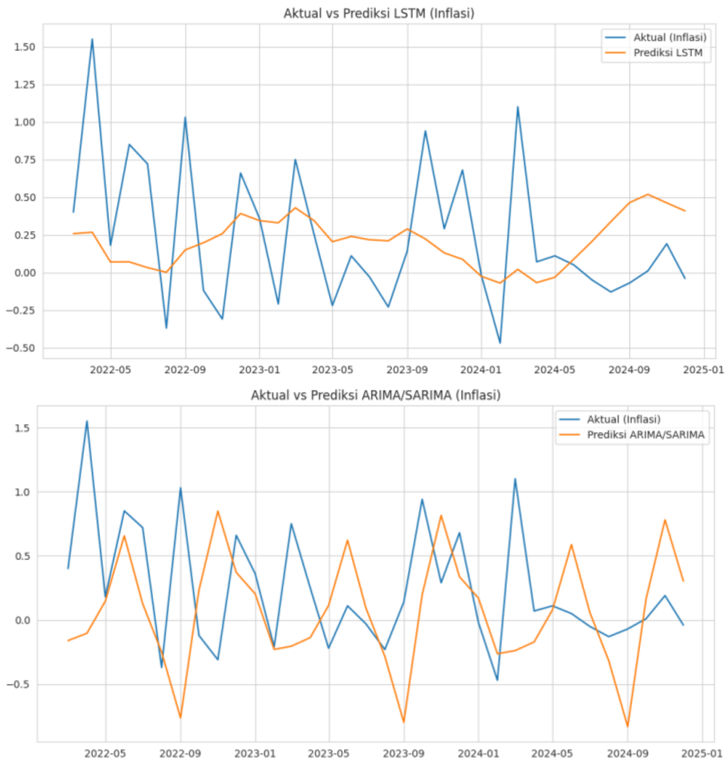

Fig. 5: Comparison of LSTM vs. SARIMA/SARIMAX predictions vs. actual values

3.6 Twelve Month Ahead Forecasts

The researcher generates 12-month-ahead forecasts for the inflation rate. For the SARIMA/SARIMAX model, projections are obtained directly by iterating the model forward. For the LSTM, future CPI scenarios (for example, held constant at the last observed value or projected separately) are constructed and then fed into the rolling input window to produce multi-step forecasts.

Table 10: Twelve-month-ahead forecasts on the test horizon

Month ahead-	SARIMA/SARIMAX	LSTM (skenario IHK)
2025+1	-0.161	0.32
2025+2	-0.104	0.283
2025+3	0.147	0.276
2025+4	0.656	0.157
2025+5	0.128	0.157
2025+6	-0.25	0.172
2025+7	-0.763	0.2
2025+8	0.231	0.234
2025+9	0.848	0.268
2025+10	0.371	0.288
2025+11	0.204	0.295
2025+12	-0.229	0.287

Table 10 and Figure 6 present the +12-month forecasts. The SARIMA/SARIMAX trajectory is generally more volatile, consistent with a seasonal linear model that is sensitive to the most recent shocks. The LSTM produces a smoother path and tends to yield small positive values over the first 12 months when the future CPI scenario is held constant at its last observed value. If the CPI scenario is itself projected (for example, using a separate SARIMA model for CPI), then the LSTM forecast path will follow that projected CPI trajectory.

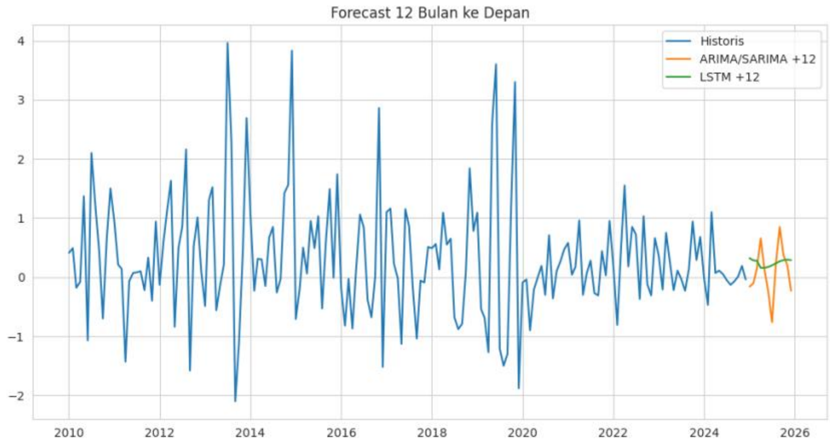

Fig. 6: Historical series vs. +12-month forecasts for SARIMA/SARIMAX and LSTM

3.7 Robustness Checks and Policy Implications

As a robustness check, we analyze sensitivity to variations in window length (6, 12, 18 months), the number of LSTM units (32–128), and a small grid of SARIMA/SARIMAX parameters. Minor changes in hyperparameters do not alter the main conclusion: the multivariate LSTM model remains superior on the primary metrics and in the Diebold–Mariano test. This indicates that the comparative result is robust.

From a policy perspective, city-level inflation forecasts can be used to prioritize price-stabilization interventions for key commodities and to support coordination of regional policy. However, the accuracy of future CPI scenarios (for the LSTM) becomes crucial; therefore, it is recommended to combine LSTM with CPI projections obtained from a separate SARIMA model or from official forecast sources.

4. Conclusion

Based on all results and analyses, the multivariate LSTM model (12-month window; Inflation and CPI as features) outperforms the SARIMA (1,1,1)×(1,1,1,12) model on the test period, as reflected by lower errors (RMSE = 0.497; MAE = 0.398) compared with SARIMA (RMSE = 0.661; MAE = 0.486). Empirically, this indicates that a modeling approach which exploits nonlinear dynamics and exogenous information (CPI) can capture monthly inflation variability more accurately for short-term horizons. The forecast trajectories are also consistent with the nature of each model: SARIMA tends to be more volatile, whereas LSTM produces smoother and more stable predictions over the first 12 months under a constant-CPI scenario.

The implication is that the choice of city-level inflation forecasting model should take into account nonlinearity, seasonality, and the availability of exogenous variables. SARIMA/SARIMAX remains a strong baseline, especially when interpretability and prediction intervals are required, but LSTM is more promising when the Inflation CPI relationship is complex or when regime changes occur. To strengthen the statistical evidence, a Diebold Mariano test on the error series of both models can be used to assess the significance of accuracy differences; nevertheless, the consistent RMSE/MAE advantage observed here already provides a practically meaningful edge for LSTM in the context of the analyzed data.

4.1. Acknowledgements

The author gratefully acknowledges all forms of support and assistance provided during this study. Appreciation is extended to colleagues and reviewers who offered constructive feedback on the study design, experimental procedures, and manuscript revisions. The author also thanks the institutional units that facilitated administrative processes, data management, and computing resources, as well as the statistical agencies and sectoral offices at the regional level that enabled access to monthly inflation and consumer price index data. Language editing and proofreading support has greatly improved the readability of this manuscript. Any financial or non-financial support (where applicable) is also sincerely acknowledged

References

[1] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 3rd ed. Melbourne, Australia: OTexts, 2021.
[2] B. Lim and Z. Zohren, "Time-series forecasting with deep learning: A survey," Philosophical Transactions of the Royal Society A, vol. 379, no. 2194, 2021, Art. no. 20200209.
[3] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "M5 accuracy competition: Results, findings and conclusions," International Journal of Forecasting, vol. 38, no. 4, pp. 1346–1364, 2022.
[4] S. Smyl, "A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting," International Journal of Forecasting, vol. 36, no. 1, pp. 75–85, 2020.
[5] P. J. Brockwell and R. A. Davis, Introduction to Time Series and Forecasting, 3rd ed. New York, NY, USA: Springer, 2016.
[6] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root," Journal of Econometrics, vol. 54, nos. 1–3, pp. 159–178, 1992.
[7] D. A. Dickey and W. A. Fuller, "Distribution of the estimators for autoregressive time series with a unit root," Journal of the American Statistical

Association, vol. 74, no. 366, pp. 427–431, 1979.
[8] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time Series Analysis: Forecasting and Control, 6th ed. Hoboken, NJ, USA: Wiley, 2015.
[9] X. Chen, S. Wei, and L. Chen, "A comprehensive survey on LSTM networks for time series forecasting," ACM Computing Surveys, vol. 55, no. 1, pp. 1–35, 2022.
[10] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," Journal of Business & Economic Statistics, vol. 13, no. 3, pp. 253–263, 1995.
[11] R. J. Hyndman and Y. Khandakar, "Automatic time series forecasting: The forecast package for R," Journal of Statistical Software, vol. 27, no. 3, pp. 1–22, 2008.
[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[13] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The M4 competition: Results, findings, conclusion and way forward," International Journal of Forecasting, vol. 34, no. 4, pp. 802–808, 2018.
[14] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," Neurocomputing, vol. 50, pp. 159–175, 2003.
[15] G. Bontempi, S. Ben Taieb, and Y.-A. Le Borgne, "Machine learning strategies for time series forecasting," European Journal of Operational Research, vol. 247, no. 3, pp. 641–651, 2015.
[16] F. Petropoulos et al., "Forecasting: theory and practice," International Journal of Forecasting, vol. 38, no. 3, pp. 705–871, 2022.
[17] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," in Proc. NeurIPS, 2020, pp. 1–13.
[18] S. N. Oreshkin, N. Carpov, N. Chapados, and Y. Bengio, "N-BEATS: Neural basis expansion analysis for interpretable time series forecasting," in Proc. ICLR, 2020, pp. 1–21.
[19] S. J. Taylor and B. Letham, "Forecasting at scale," The American Statistician, vol. 72, no. 1, pp. 37–45, 2018.
[20] D. I. Harvey, S. J. Leybourne, and P. Newbold, "Testing the equality of prediction mean squared errors," International Journal of Forecasting, vol. 13, no. 2, pp. 281–291, 1997