Forecasting Consumer Price Index of Malaysia: A Comparative Study of Machine & Deep Learning Approaches

Table of Content

List of Table

List of Figure

1 Introduction

The **Consumer Price Index (CPI)** is a critical economic indicator used to measure inflation and guide monetary policy, income adjustments, and business planning. In Malaysia, CPI forecasting supports government initiatives, business strategies, and household financial decisions. However, most CPI forecasts in Malaysia rely on traditional statistical models, with limited use of modern **machine learning (ML)** or **deep learning (DL)** methods.

This project addresses key research gaps by comparing the performance of four models—**SVR, Random Forest, XGBoost, and Neural Networks**—across three input structures: univariate (CPI only), multivariate (CPI + 12 indicators), and simple multivariate (CPI + 3 top indicators). It also tests the best-performing model's adaptability to **Cambodia, Myanmar, and Laos**, providing a broader regional perspective for CPI forecasting in Southeast Asia.

2 Result

2.1 Univariate Input *(CPI Only)*

Model	MAE	MSE	RMSE	R2
SVR	0.644567	0.859446	0.927063	0.996574
RF	0.389592	0.385025	0.620504	0.998465
XGBoost	0.420826	0.36499	0.604144	0.998545
NN	0.775079	1.484162	1.218262	0.994083

Table 1 Result Univariate Input.

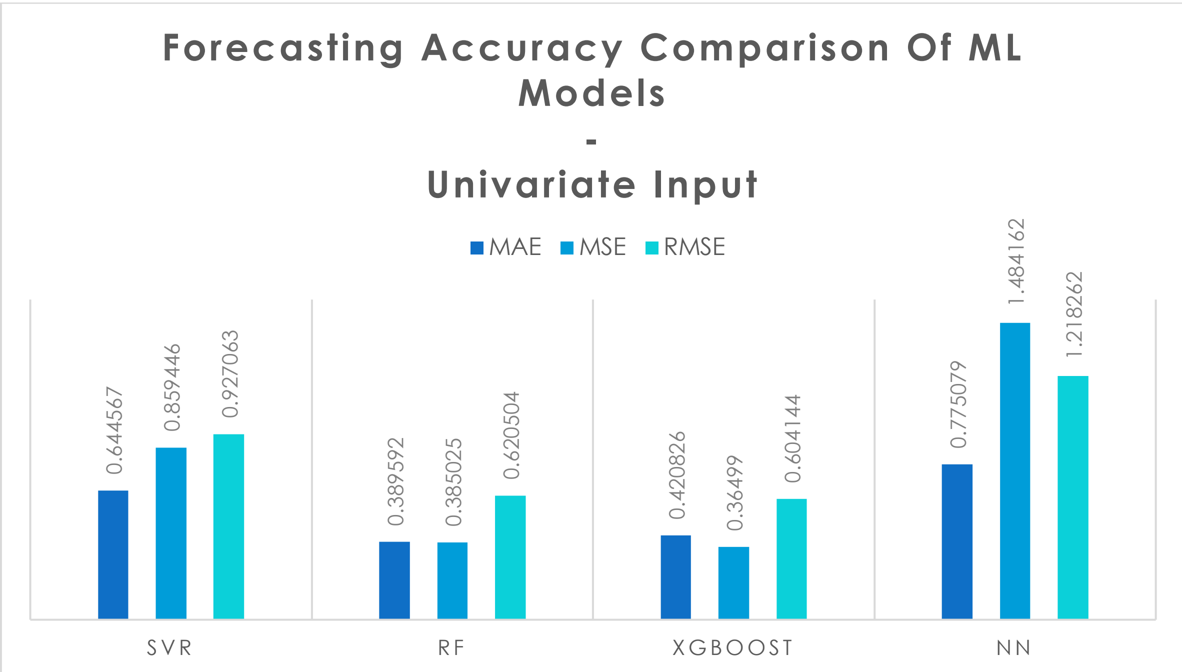

Figure 1 Forecasting Accuracy Comparison Of ML Models- Univariate Input

When using only the CPI as the input feature, all four models demonstrated strong predictive capabilities, as reflected in their high R^2 scores (*Table 1*). Among them, *RF* delivered the best overall performance, achieving the lowest MAE of 0.3896 and the lowest MSE of 0.3850, indicating high predictive accuracy and minimal error. *XGBoost* followed closely, with the lowest RMSE of 0.6041, showcasing effective handling of smaller prediction errors. *SVR* achieved a solid R^2 score of 0.9966, but its higher MAE (0.6446) and RMSE (0.9271) suggested it was less efficient in capturing CPI trends under the univariate structure. *NN* model exhibited the weakest performance among the four, with the highest error values (MAE = 0.7751, MSE = 1.4842) and the lowest R^2 score (0.9941), which may be attributed to the model's sensitivity to data volume and parameter tuning.

2.2 Multivariate Input (12 Economic Indicators)

Model	MAE	MSE	RMSE	R2
SVR	0.058355	0.005041	0.071001	0.999973
RF	0.243706	0.193926	0.440371	0.998955
XGBoost	0.271007	0.147739	0.384368	0.999204
NN	1.058589	2.12154	1.456551	0.988567

Table 2 Result Multivariate Input (12 Economic Indicators)

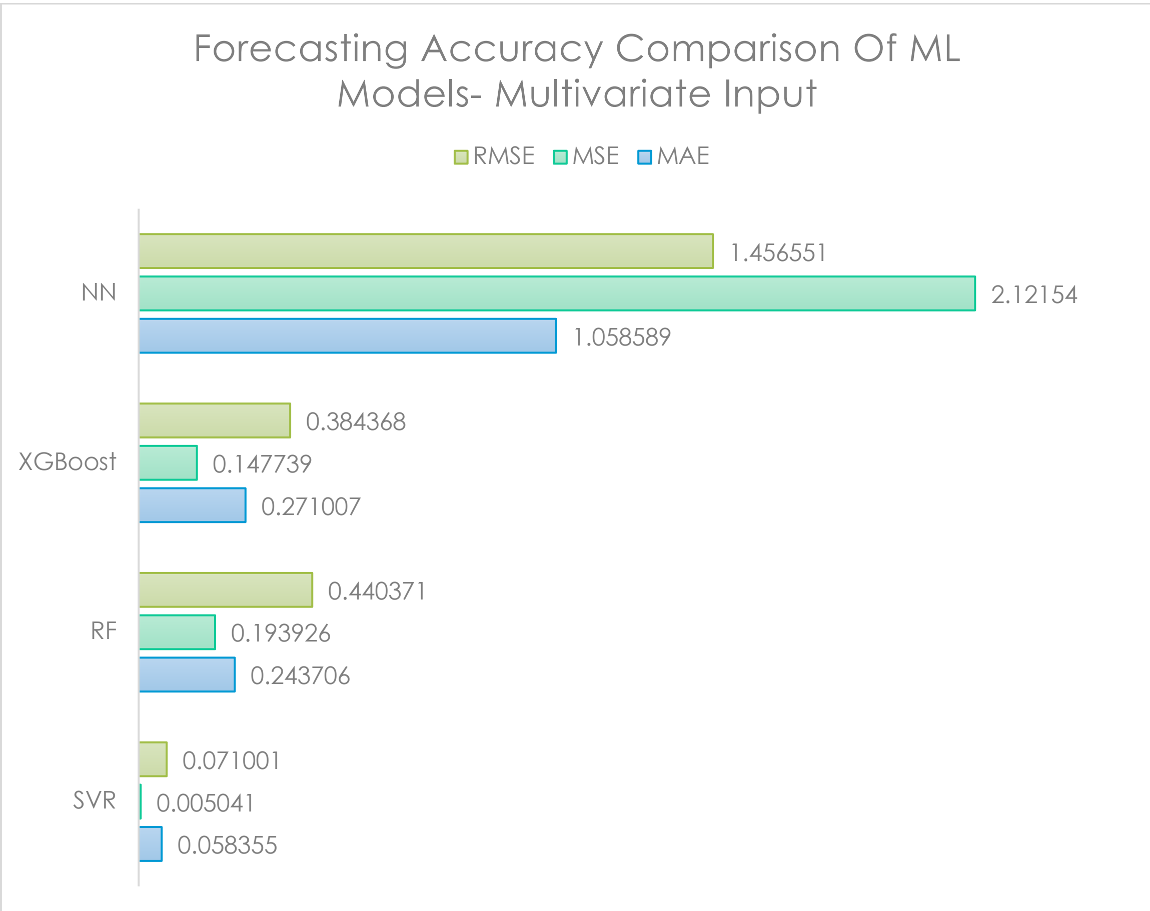

Figure 2 Forecasting Accuracy Comparison Of ML Models- Multivariate Input

When all 12 economic indicators were included as input features, the performance of all models improved significantly, highlighting the value of multivariate data in enhancing CPI forecasting accuracy, as summarized in the *Table 2*. S*VR* emerged as the top-performing model, recording the lowest MAE (0.0584) and RMSE (0.0710), along with a near-perfect R^2 score of 0.99997—indicating an exceptional fit and the model's ability to capture complex relationships between indicators and CPI trends. *XGBoost* and *RF* also performed robustly, with R^2 scores exceeding 0.998 and substantially lower error metrics compared to their univariate counterparts, demonstrating their strength in handling diverse input features. Although the *NN* remained the lowest-performing model in this configuration, it showed modest improvements in both error metrics and R^2 score, suggesting some benefit from the richer feature set, though it continued to lag the other approaches.

2.3 Simple Multivariate Input (3 Most Correlated Economic Indicators)

Model	MAE	MSE	RMSE	R2
SVR	0.380881	0.434028	0.658808	0.997661
RF	0.326505	0.393973	0.627673	0.997877
XGBoost	0.402543	0.432903	0.657954	0.997667
NN	0.684736	0.786823	0.88703	0.99576

Table 3 Result Multivariate Input (3 Correlated Economic Indicators)

Figure 3 Forecasting Accuracy Comparison Of ML Models – Simple Multivariate Input

To optimize model efficiency and reduce computational complexity, the dataset was further refined to include only the three features most strongly correlated with the CPI. While this simplification led to a moderate decline in performance compared to the 12-feature configuration, the results remained robust (*Table* 3). *RF* delivered the best overall performance in this setup, with a MAE of 0.3265 and RMSE of 0.6277, reflecting strong predictive accuracy despite the reduced input. *SVR* and *XGBoost* also performed well, each maintaining R^2 scores above 0.997, indicating that the models were still able to capture most of the variance in the data. *NN* continued to underperform relative to the other models, with an MAE of 0.6847 and an R^2 of 0.9958. Although slightly less accurate than the full-feature model, this streamlined structure struck a good balance between model simplicity and forecasting performance.

2.4 Cross-Country Generalizability of SVR (12 Indicators)

Model	Country	MAE	MSE	RMSE	R^2 Score
SVR	**Cambodia**	0.222801	0.067534	0.259874	0.999958
	Myanmar	0.120868	0.024862	0.157676	0.999928
	Lao People's Dem. Rep.	0.188276	0.045322	0.212889	0.999973
	Malaysia	0.058355	0.005041	0.071001	0.999973

Table 4 Result Cross-Country

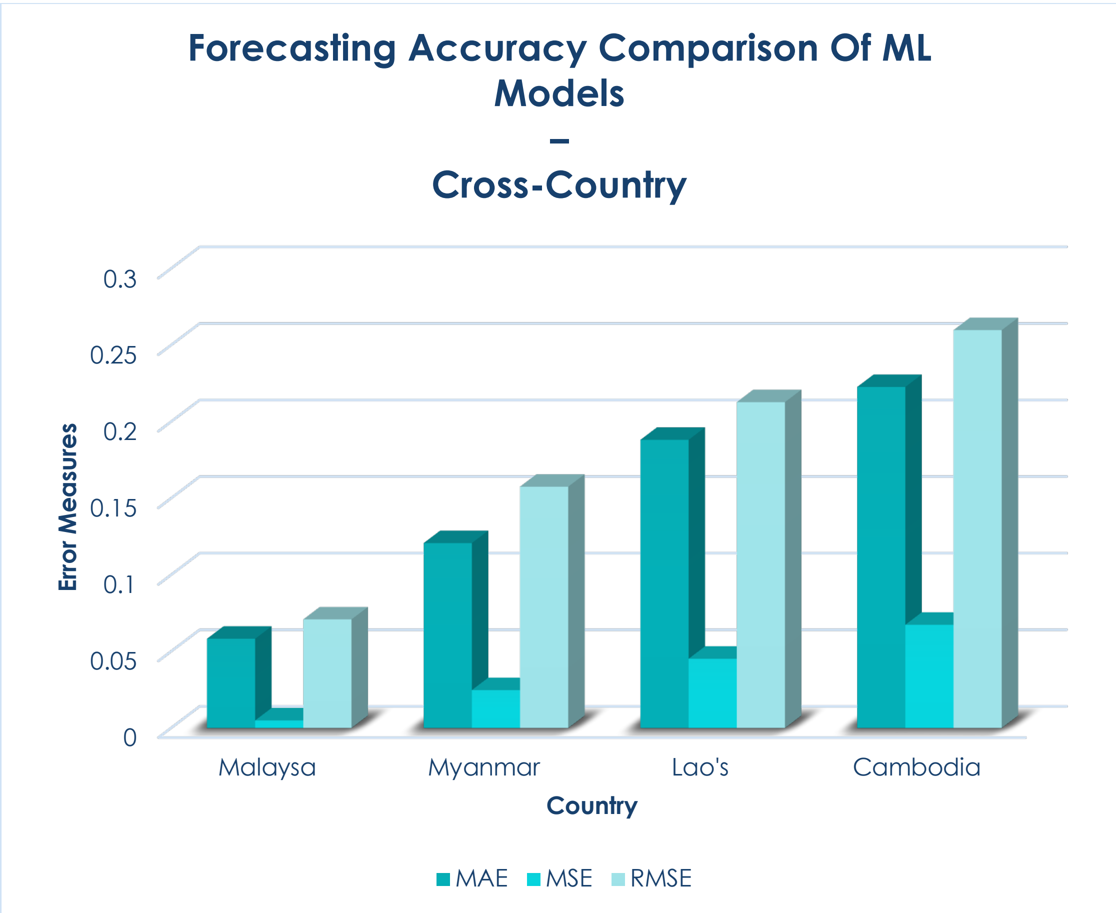

Figure 4 Forecasting Accuracy Comparison Of ML Models – Cross-Country

To assess the generalizability of the best-performing model, the *SVR* with 12 economic indicators was applied to three additional Southeast Asian countries:

Cambodia, Myanmar, and Laos. As detailed in *Table 4* the model consistently demonstrated robust predictive performance across all countries. In Cambodia, the SVR achieved an MAE of 0.2228, RMSE of 0.2599, and an R^2 score of 0.99996. Similarly, in Myanmar, the model recorded an MAE of 0.1209, RMSE of 0.1577, and R^2 of 0.99993. For Laos, the results remained strong with an MAE of 0.1883, RMSE of 0.2129, and R^2 of 0.99997. Comparatively, the original performance in Malaysia yielded the lowest errors, with an MAE of 0.0584 and RMSE of 0.0710, and an equally impressive R^2 of 0.99997. These outcomes confirm that SVR not only fits Malaysia's dataset exceptionally well but also generalizes effectively across different economic environments, maintaining high accuracy and low error rates.