Hybrid Time Series Forecasting of USD/INR Using ARIMA and XGBoost

Author: Hanumantha Sai Karthik Velavarthypathi

Abstract

This project presents a hybrid approach to forecasting the USD/INR exchange rate by combining traditional time series modeling (ARIMA) with modern machine learning (XGBoost). While ARIMA effectively captures linear temporal patterns, it fails to model nonlinear fluctuations that often arise in financial time series. To address this, we model ARIMA residuals using XGBoost, enabling the system to learn complex nonlinear structures. Our hybrid model demonstrates a significant reduction in forecasting error, achieving a Mean Squared Error (MSE) of 0.0057 compared to 0.1001 for standalone ARIMA. The results underscore the power of combining statistical and machine learning techniques for improved accuracy in financial forecasting.

Problem Statement

Accurate currency forecasting is critical for financial institutions, businesses, and investors exposed to foreign exchange (Forex) market risk. Traditional models like ARIMA, while statistically sound, often underperform in capturing the nonlinear and volatile nature of exchange rates. This project aims to enhance forecasting accuracy for the USD/INR exchange rate by:

- Building a traditional ARIMA model for baseline forecasting.
- Modeling the residual errors from ARIMA using a machine learning model (XGBoost).
- Combining both predictions to form a hybrid model with improved performance.

Methodology

1. Data Preprocessing

- Historical USD/INR exchange rate data (approx. 6,000 business days) was cleaned.

Hybrid Time Series Forecasting of USD/INR Using ARIMA and XGBoost

- Duplicate dates were removed, and the series was resampled to a business-day frequency.

- Stationarity was tested using the Augmented Dickey-Fuller test. First-order differencing made the series stationary (d=1).

2. ARIMA Modeling

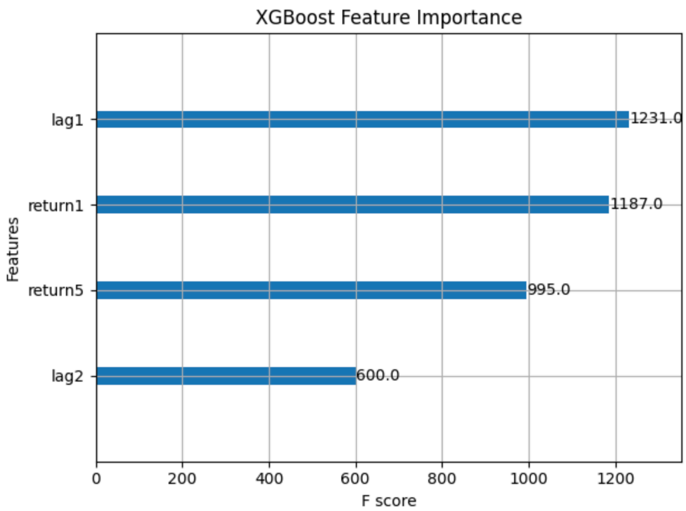
- auto_arima() suggested an ARIMA(5,1,3) model.

- The model was trained on the full time series.

- Fitted values were extracted and used to calculate residuals.

3. Residual Modeling with XGBoost

- Features engineered:
  - Lagged exchange rates (lag1, lag2)
  - Short- and medium-term returns (return1, return5)
- XGBoost was trained on these features to predict residuals.
- Final hybrid forecast = ARIMA prediction + ML-predicted residual.



4. Evaluation

- Last 30 valid data points were held out for backtesting.

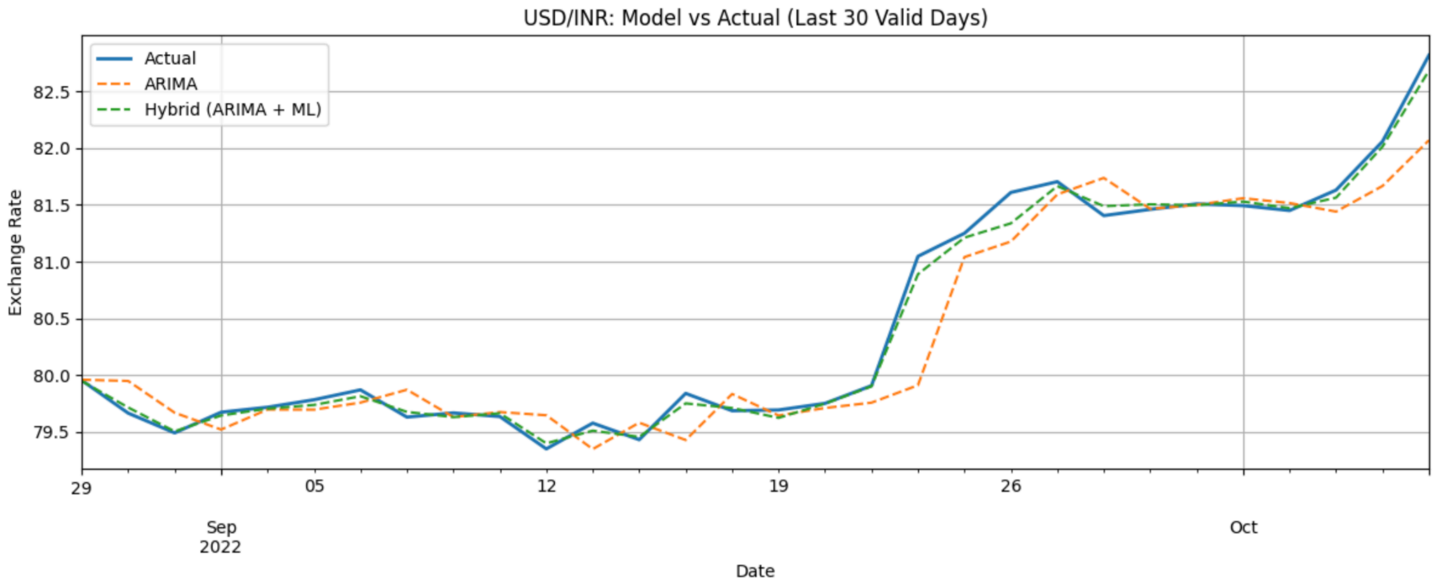
- MSE was used as the evaluation metric for comparison.

Hybrid Time Series Forecasting of USD/INR Using ARIMA and XGBoost

Results

ARIMA MSE: 0.100091
Hybrid MSE: 0.00571

- The hybrid model reduced the error by over 90%, confirming its effectiveness.
- Feature importance from XGBoost showed that lag1, return1, and return5 were key drivers of residual behavior.
- Visual comparisons showed the hybrid forecast closely tracked actual values, especially during short-term fluctuations.



Conclusion

This project validates the effectiveness of a hybrid ARIMA + ML approach for time series forecasting in finance. By combining ARIMA's strength in modeling linear dependencies with XGBoost's ability to learn nonlinear residual structures, the hybrid model significantly outperforms ARIMA alone. Such models can be adapted to other financial time series (stocks, commodities, crypto) and embedded into real-time trading, risk management, or decision Support tools