Short Report: Stock Price Forecasting (DataSynthis ML Job Task)

Overview:

The objective of this work was to implement stock price forecasting using both traditional statistical and machine learning methods and evaluate their performance under a rolling window evaluation framework. This approach simulates real-world forecasting situations where models are continuously retrained on the most recent data before predicting the next move.

Methods:

• **ARIMA (5,1,0):** A classical time-series model suitable for capturing short-term linear dependence and trends. The model was updated iteratively using a rolling window to simulate dynamic market conditions.

• **LSTM (Long Short-Term Memory):** A recurrent neural network capable of modeling long-range dependence and nonlinear behavior in sequential data. It was trained on normalized price data with a 10-day input sequence to predict the next closing price.

Evaluation Metrics:

• **RMSE (Root Mean Square Error):** To measure overall predictive accuracy.

• **MAPE (Mean Absolute Percentage Error):** To capture percentage-based deviation, making relative error easier to interpret.

Results:

Model	RMSE (%)	MAPE (%)	Accuracy (%)
ARIMA	6.633162	1.629443	98.370557
LSTM	13.462030	4.125158	95.874842

Discussion:

• ARIMA has shown stable results and is computationally efficient. It is interpretable and effective in predicting short-term linear trends, but struggles with volatility and sudden regime changes in stock data.

• LSTM captures more complex nonlinear patterns, making it better suited for financial time-series where price movements are influenced by multiple interacting factors. However, it requires more training time, tuning, and sufficient data to achieve good results.

• Rolling window evaluations highlight that LSTM adapts better to changing patterns over time, while ARIMA's performance degrades when the data exhibits higher volatility.

Conclusion:

Both ARIMA and LSTM are valuable methods for time-series forecasting. ARIMA serves as a strong baseline due to its simplicity and interpretability, while LSTM offers superior generalization capabilities, especially in capturing the nonlinear dynamics present in the stock market. For real-world applications in DataSynthis, LSTM-based models will be more convenient for long-term deployment, while ARIMA can be retained for rapid benchmarking and validation.