From ARIMA to LSTM: A Comparative Study in Stock Price Forecasting

Predicting the stock market is one of the most challenging and intriguing problems in the world of finance and data science. The volatile, non-linear nature of stock prices makes it a perfect testing ground for a wide range of forecasting models. In this project, we took on this challenge, using historical data from Tesla (TSLA) to forecast future stock prices.

Our goal was not just to build a single model, but to compare two fundamentally different approaches: a classic statistical method, **ARIMA**, and a powerful deep learning model, **LSTM**.

The Two Contenders: ARIMA vs. LSTM

1. ARIMA (AutoRegressive Integrated Moving Average)

ARIMA is a powerhouse in traditional time-series forecasting. It works by analyzing the statistical properties of past data to predict future values. It's a linear model, meaning it assumes that the relationships in the data can be described by a straight line. ARIMA is great for capturing trends and seasonality, and it serves as an excellent baseline model.

2. LSTM (Long Short-Term Memory)

LSTM is a special type of Recurrent Neural Network (RNN) designed specifically for sequence data, like time series. Unlike ARIMA, LSTMs are non-linear models. Their key feature is their "memory," which allows them to remember important patterns over long periods while forgetting irrelevant noise. This makes them exceptionally well-suited for understanding the complex, long-term dependencies that drive stock market behavior.

The Experiment: Rolling Window Evaluation

To rigorously test our models, we used a **rolling window evaluation**. Instead of a single train-test split, we trained and tested our models on multiple, overlapping "windows" of data. This method simulates how a model would perform in the real world, where it would need to be retrained periodically as new data becomes available.

For each window, we measured two key metrics:

- **RMSE (Root Mean Squared Error)**: Measures the average magnitude of the prediction errors. A lower RMSE is better.
- **MAPE (Mean Absolute Percentage Error)**: Measures the average percentage error,

making it easy to interpret. A lower MAPE is better.

The Results: A Clear Winner Emerges

After running both models through the rolling window evaluation, the results were conclusive. The **LSTM model consistently outperformed the ARIMA model**.

Model	Average RMSE	Average MAPE (%)
LSTM	**81.85**	**12.82%**
ARIMA	129.39	19.34%

As the table shows, the LSTM model had a significantly lower average RMSE and MAPE, indicating that its predictions were more accurate and reliable across all the test windows.

Why Did LSTM Win?

The stock market, especially for a volatile stock like TSLA, is anything but linear. The LSTM's architecture is inherently designed to capture the complex, non-linear relationships and long-term patterns that ARIMA, as a linear model, simply cannot see. The LSTM's ability to "remember" past price movements and their impact over time gives it a decisive edge in this kind of forecasting task.

From Analysis to Application: Deployment

To bring this project to life, the superior LSTM model was deployed as an interactive web application on **Hugging Face Spaces** using **Gradio**. This final step transforms the project from a research notebook into a practical tool that anyone can use to get real-time stock price forecasts.

Conclusion

This project demonstrates the power of deep learning in modern time-series forecasting. While traditional models like ARIMA are still valuable, the superior performance of the LSTM highlights the importance of choosing the right tool for the job. For complex, dynamic datasets like stock prices, the ability of neural networks to learn intricate patterns makes them an invaluable asset for any data scientist.