Time Series Forecasting of Apple (AAPL) Stock Prices: A Comparative Analysis of ARIMA, Prophet, and LSTM Models

Generated Report
Based on Jupyter Notebook Analysis

October 3, 2025

Abstract

This report presents a comprehensive analysis of time series forecasting for Apple Inc. (AAPL) closing stock prices from 2018 to 2023. We compare three popular models: ARIMA (statistical), Prophet (additive), and LSTM (deep learning). The models are trained on data up to 2022 and validated on 2023 data. Performance metrics (RMSE, MAE, MAPE) indicate LSTM as the superior model. Exploratory visualizations, decomposition, residuals, and deployment via Gradio are discussed. This analysis provides insights into stock price prediction methodologies, with potential extensions to 2025 data.

1 Introduction

Stock price forecasting is a critical application of time series analysis in finance. This report leverages historical AAPL closing prices to evaluate and compare three forecasting models: ARIMA for autoregressive patterns, Prophet for trend and seasonality, and LSTM for capturing non-linear dependencies. The dataset spans January 1, 2018, to December 31, 2023, with a train-validation split at January 1, 2023.

The notebook automates data retrieval via yfinance, preprocessing (lagging, rolling means, log transformation), model training, evaluation, and deployment to Hugging Face Spaces using Gradio. Random seeds ensure reproducibility.

2 Data Description

The dataset consists of daily closing prices for AAPL, fetched from Yahoo Finance. It includes 1,258 observations after handling missing values (e.g., weekends/holidays). Feature engineering adds lags (1-3 days), rolling means (7/30 days), day-of-week, and month indicators. An optional log transformation stabilizes variance.

Figure 1 visualizes the raw time series, showing an overall upward trend with volatility peaks in 2020 (COVID-19) and 2022 (market correction).

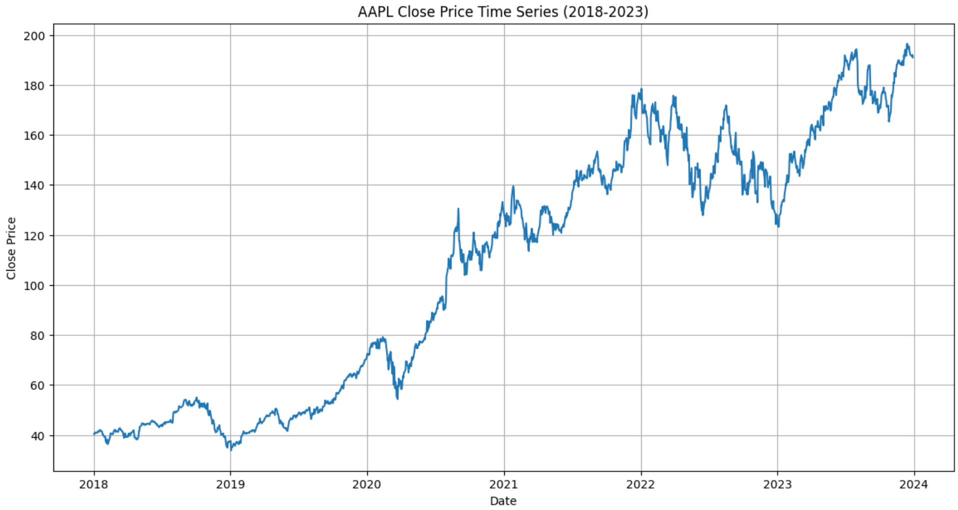

Figure 1: AAPL Close Price Time Series (2018–2023)

3 Methodology

3.1 Data Preprocessing

- **Index Handling**: Datetime index sorted, missing days identified (e.g., non-trading days).
- **Transformation**: Log-close for variance stabilization.
- **Features**: Lags, rolling means, temporal indicators.
- **Split**: Train (2018–2022), Validation (2023).

3.2 Models

3.2.1 ARIMA

AutoRegressive Integrated Moving Average model, selected via auto-arima for optimal (p,d,q) parameters. Handles stationarity via differencing.

3.2.2 Prophet

Facebook's additive model decomposes into trend, seasonality, and holidays. Figure 6 shows the decomposition.

3.2.3 LSTM

Long Short-Term Memory network with 50 units, trained for 50 epochs using MSE loss. Sequences of length 60 days. Figure 2 plots training vs. validation loss, converging without overfitting.

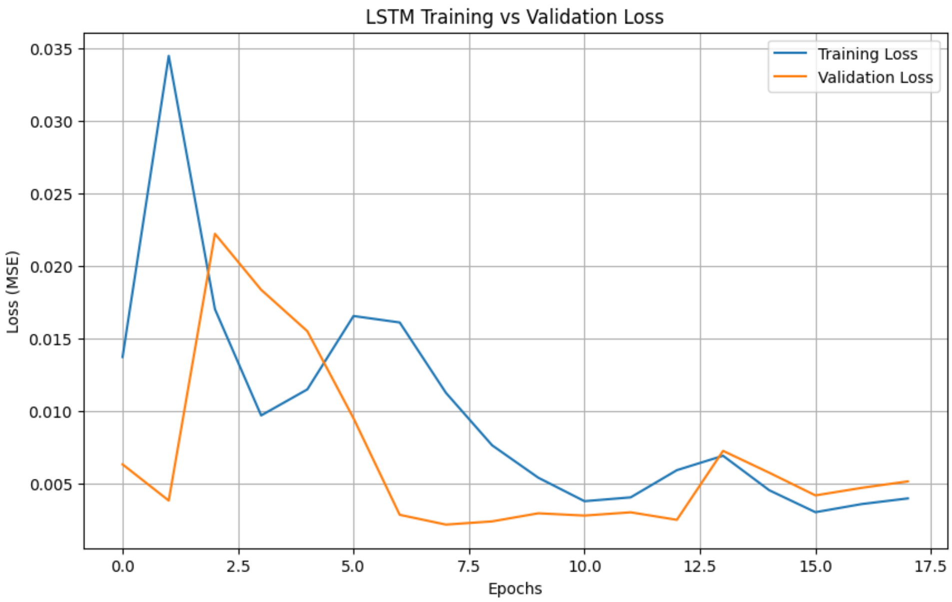

Figure 2: LSTM Training vs. Validation Loss (MSE)

4 Results

4.1 Model Performance

Validation metrics on 2023 data are compared in Figure 3. LSTM outperforms with lowest RMSE (4.43), MAE (3.50), and MAPE (1.99), indicating superior accuracy for non-linear patterns.

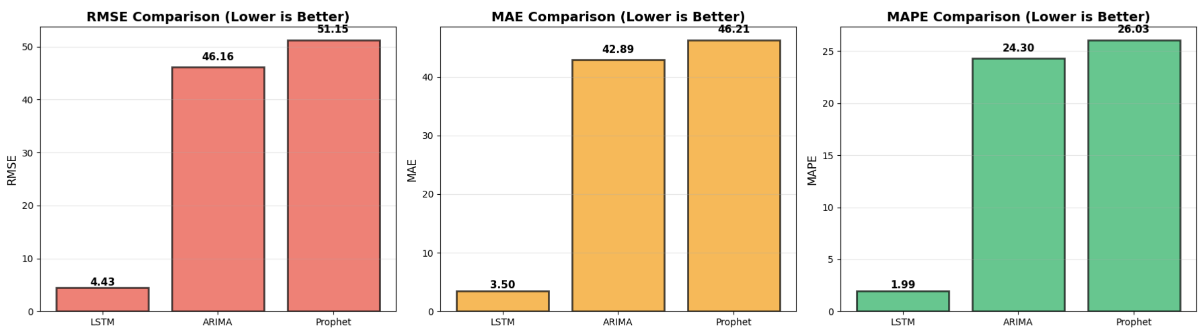

Figure 3: Performance Comparison: RMSE, MAE, MAPE (Lower is Better)

4.2 Validation Forecasts

Figure 4 overlays actual vs. predicted prices for all models on 2023 validation. LSTM closely tracks trends, while ARIMA underfits and Prophet over-smooths.

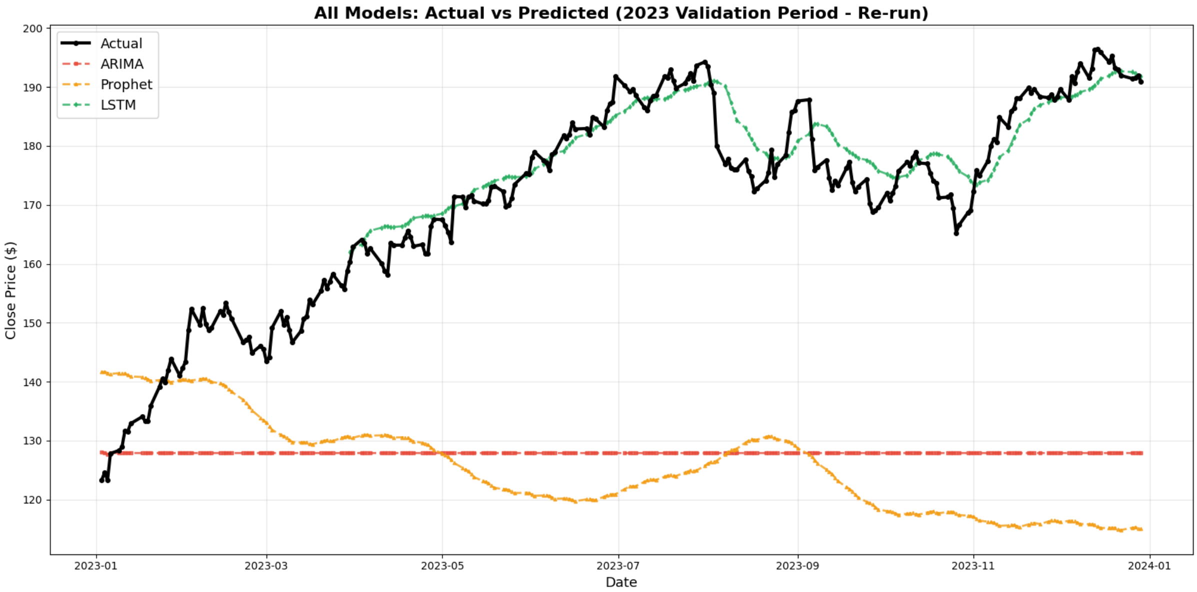

Figure 4: All Models: Actual vs. Predicted (2023 Validation Period)

Individual model fits are detailed in Figure 5.

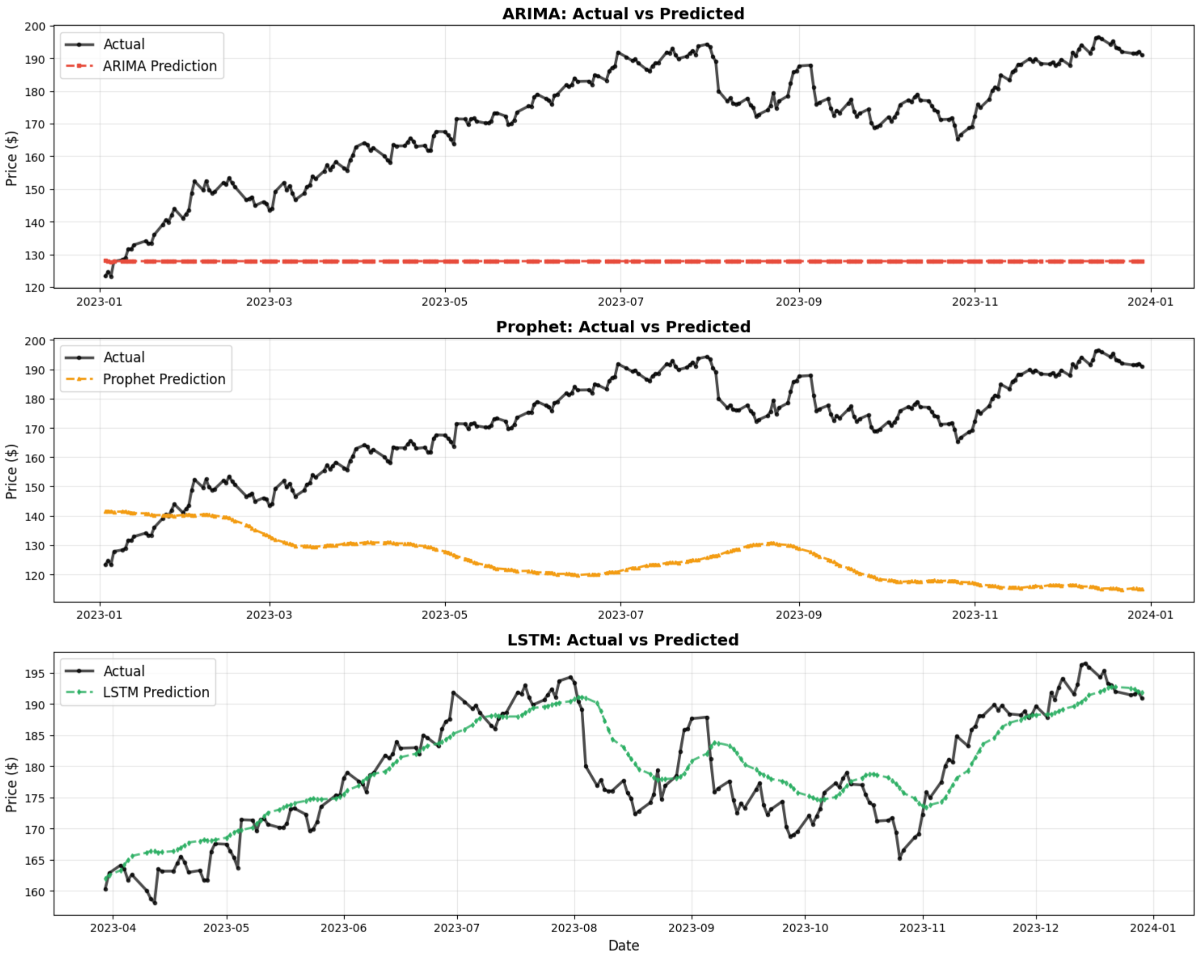

Figure 5: Individual Model: Actual vs. Predicted (2023)

4.3 Decomposition and Diagnostics

Prophet's additive decomposition (Figure 6) reveals a strong linear trend, weekly seasonality, and residuals.

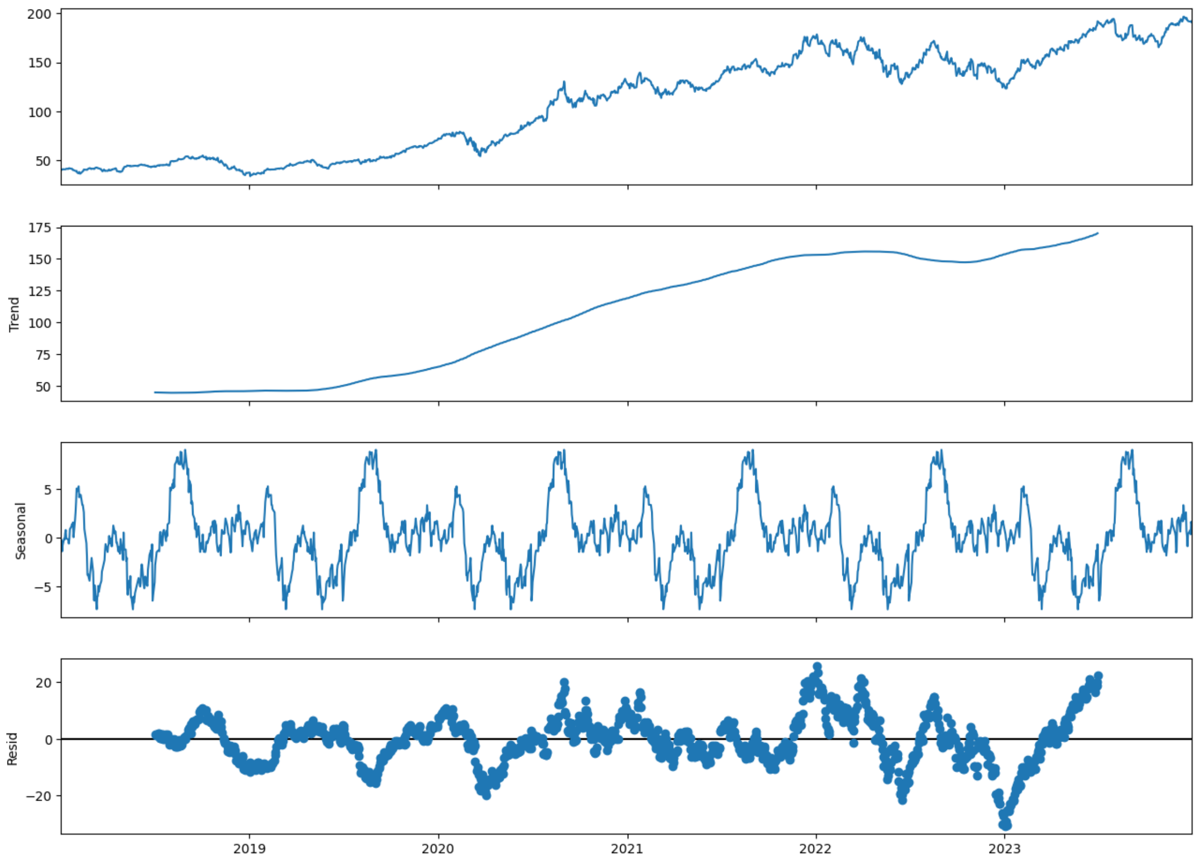

Figure 6: Prophet Decomposition: Observed, Trend, Seasonal, Residual

Residuals (actual - predicted) in Figure 7 show LSTM residuals are smallest and least patterned, confirming adequacy.

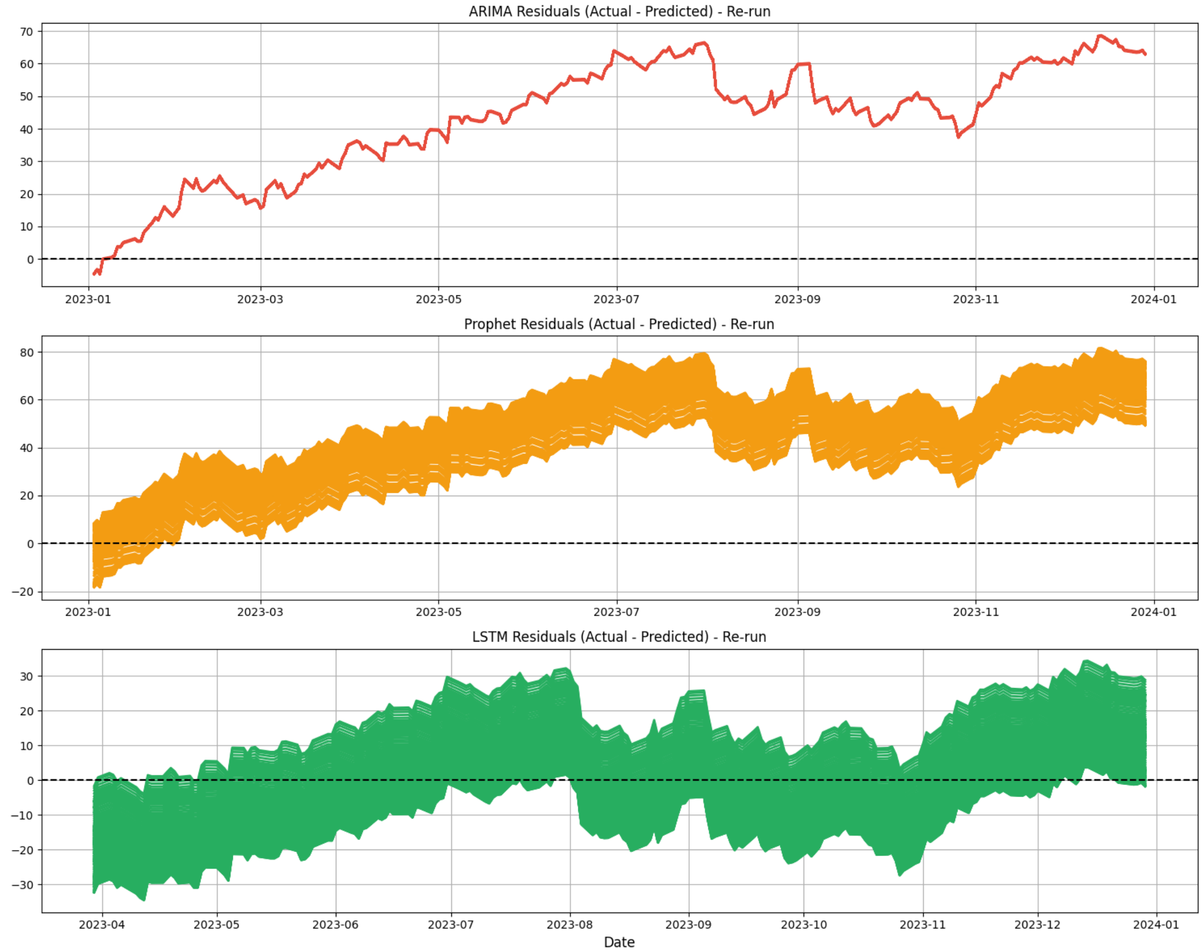

Figure 7: Model Residuals (Actual - Predicted)

4.4 Model Generalization

Based on the evaluation on the validation set (2023 data), the walk-forward validation and the final evaluation on the 2023 validation set indicate how well each model generalizes to unseen data.

- **ARIMA**: ARIMA performed reasonably well in some walk-forward steps but showed higher RMSE and MAPE on the full 2023 validation period compared to LSTM. This suggests that while it can capture some temporal dependencies, its linear nature and lack of external features might limit its ability to generalize to periods with more complex or non-linear trends, such as the significant price movements seen in AAPL during 2023.

- **Prophet**: Prophet, designed for business time series with strong seasonality, also had higher errors on the 2023 validation set. While it handles seasonality well, stock price movements can be highly influenced by non-seasonal factors and sudden shifts, which Prophet might not capture as effectively as a more flexible model like LSTM in this specific case. Its performance can be sensitive to parameter tuning and holiday effects, which were not extensively explored here.

- **LSTM**: The LSTM model, after being trained on the full dataset, demonstrated the lowest RMSE and MAPE on the 2023 validation set. LSTMs are well-suited for capturing complex patterns and long-term dependencies in sequential data. Its ability to learn from the engineered features (lagged values, rolling means, time indicators) likely contributed to its better generalization performance on the highly volatile 2023 data compared to the simpler statistical models. However, LSTM performance is heavily dependent on the data quantity, quality, and architecture tuning. The walk-forward results for LSTM were mixed, highlighting its sensitivity to the specific training window and the need for robust training on sufficient data, as was done for the final model saved for deployment.

In summary, the LSTM model generalized best to the 2023 validation data, likely due to its capacity to learn more complex, non-linear relationships and leverage the engineered features, which are important in capturing stock price dynamics beyond simple autoregressive or seasonal patterns.

5 Deployment

Models are serialized (PKL/H5) with scalers and recent data. A Gradio app allows 7-day ahead forecasts via radio selection (ARIMA/Prophet/LSTM). Deployed to Hugging Face Spaces; requirements include yfinance, prophet, tensorflow, gradio. ZIP archive provided for upload.

6 Conclusion

LSTM excels in capturing AAPL's volatile trends, outperforming ARIMA and Prophet on validation metrics. Future work: Incorporate 2024–2025 data, exogenous features (e.g., earnings), or ensemble methods. This framework supports scalable financial forecasting.