Long-Term Forecasting with Hybrid Time Series Models for Financial Data

1. Introduction

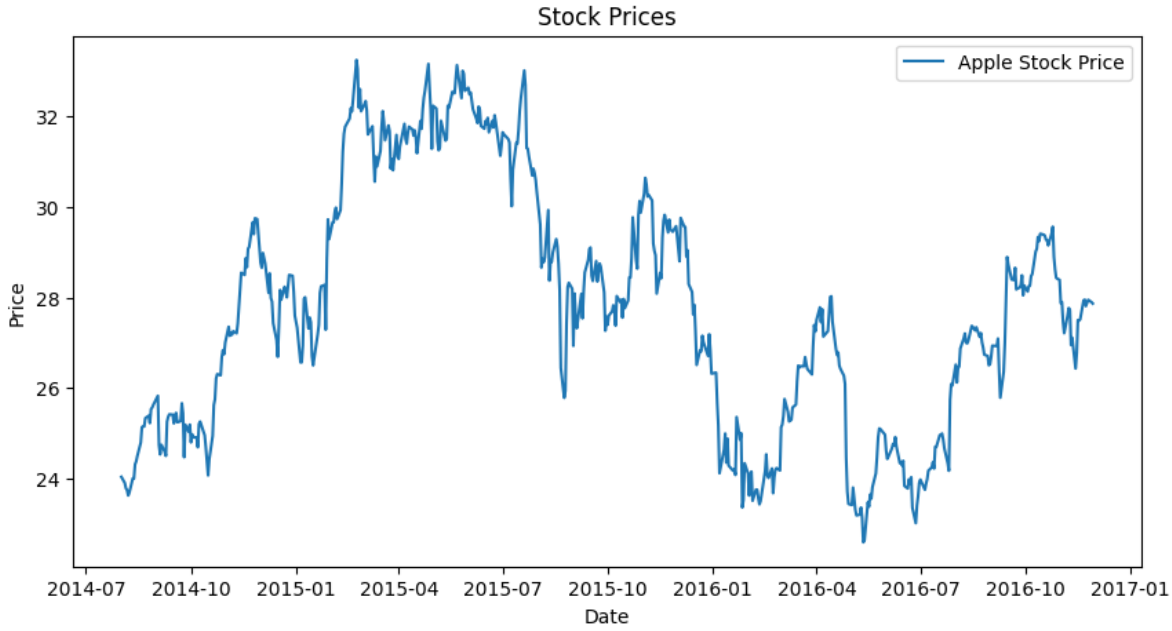
The goal of this project is to predict the stock price of Apple Inc. (AAPL) using historical stock market data. The project explores statistical time series modelling (ARIMA) and deep learning approaches (LSTM), culminating in a proposed hybrid model to capture both linear and non-linear trends.

2. Data Collection

- Source: Yahoo Finance (via the yfinance library) – AAPL ticker – Downloaded .csv format

Visualizations:

1. Raw Data Visualization: The time series of AAPL closing prices was plotted to identify general trends.



3. Data Preprocessing

Steps:

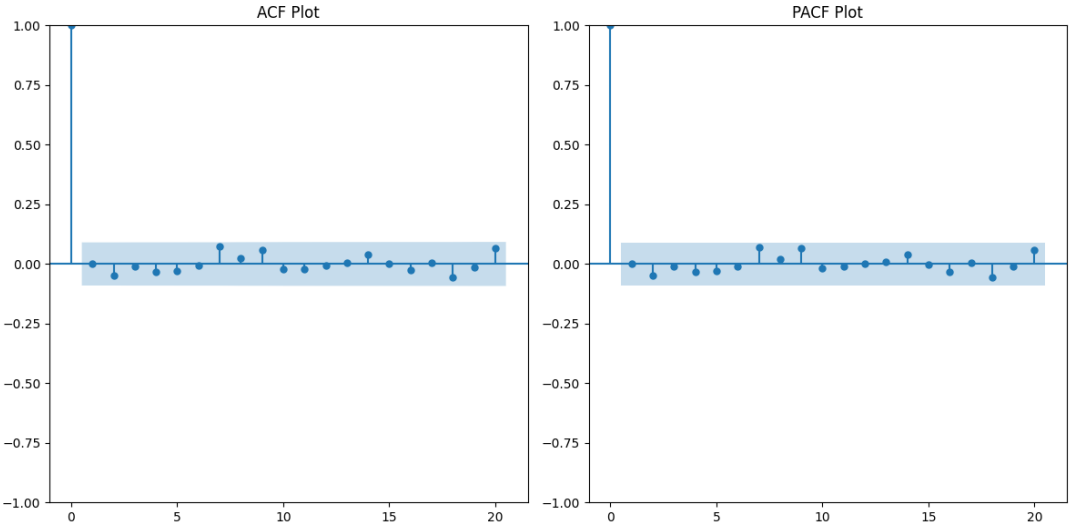
1. Handling Missing Data:
   - No missing values were present in the dataset.
2. Stationarity Test:
   - The Augmented Dickey-Fuller (ADF) test was conducted. Results:
     - ADF Statistic: -2.1682503299850526
     - p-value: 0.2179698405819862
     - The series is non-stationary. Differencing will be applied.

3. Differencing:

- First-order differencing applied; ADF confirmed stationarity after differencing.

4. ACF/PACF Plots:

- Autocorrelation (ACF) and Partial Autocorrelation (PACF) plots were generated to identify the ARIMA model's p and q parameters.



4. Feature Engineering

To improve model performance, the following features were added:

1. Moving Averages:
   - 20-day and 50-day moving averages to capture medium-term trends.
2. Lag Features:
   - Lagged values of 1 and 2 days to capture short-term dependencies.

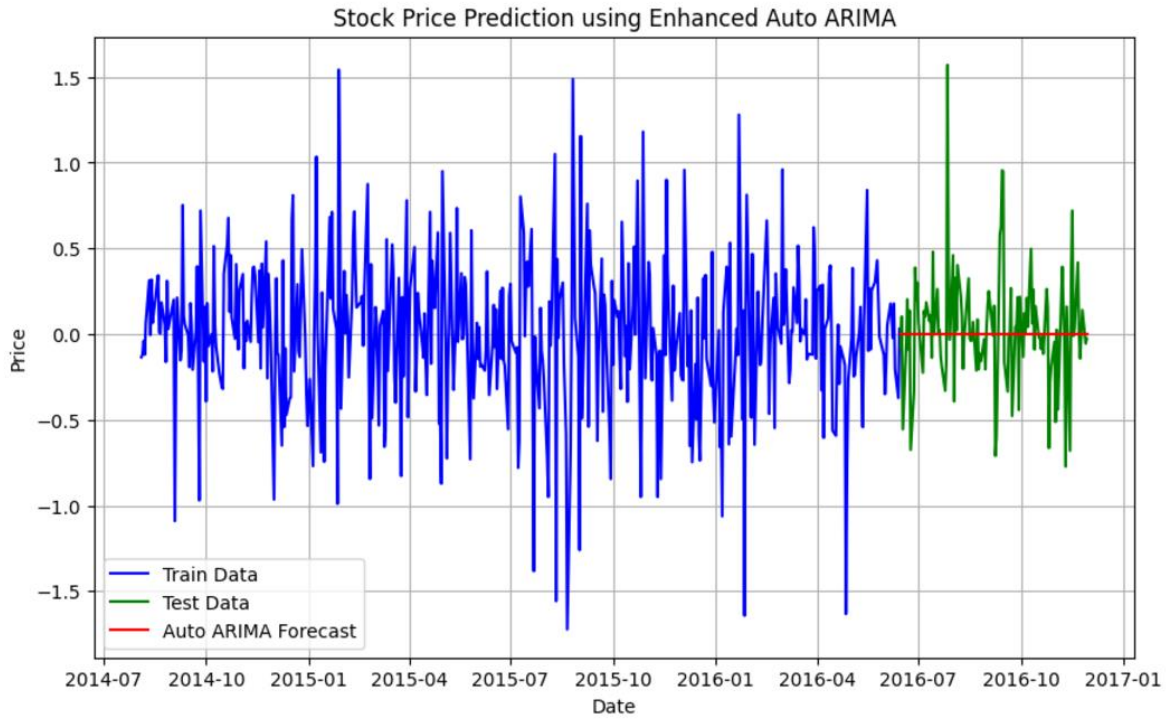
5. Statistical Modeling: Auto ARIMA

Model Selection:

- Search Space:
  - ARIMA parameters: p, d, q = (0-5)
  - Seasonal Parameters: P, D, Q, m = (0-3, m = 12 months)
  - Best Model: SARIMAX(0, 0, 0)(0, 0, 0)[12]
- Auto ARIMA suggested no significant seasonal or autoregressive components were necessary.

Model Training and Evaluation:

- The dataset was split into 80% training and 20% testing data.
- Evaluation Metrics:
  - Mean Absolute Error (MAE): 0.23479
  - Root Mean Squared Error (RMSE): 0.33983



6. Proposed LSTM Implementation

1. LSTM Model Architecture

Layers:

- Bidirectional LSTM: 50 units with ReLU activation, processing data in both forward and backward directions.
- Dropout Layer: 20% dropout for regularization to prevent overfitting.
- Second Bidirectional LSTM: Another 50-unit layer for deeper temporal modeling.
- Dropout Layer: 20% dropout for additional regularization.
- Dense Layer: A fully connected layer with 1 unit for the final prediction.

Compilation:

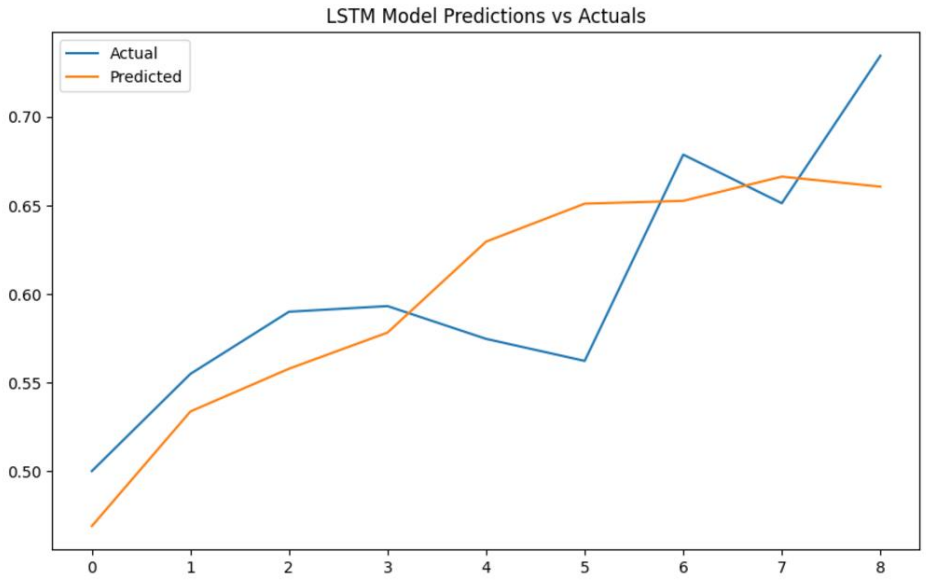
- Optimizer: Adam for adaptive learning rate.
- Loss Function: Mean Squared Error (MSE) to minimize prediction error.

2. Model Training

- Training Settings:
- Epochs: Up to 100.
- Batch Size: 32 sequences processed in parallel.
- Validation Split: 20% of training data reserved for validation.
- Early Stopping: Monitors validation loss and halts training after 10 epochs without improvement, restoring the best weights.

3. Model Evaluation

- Performance Metrics:
  - Root Mean Squared Error (RMSE): 0.0471
  - Mean Absolute Error (MAE): 0.0398



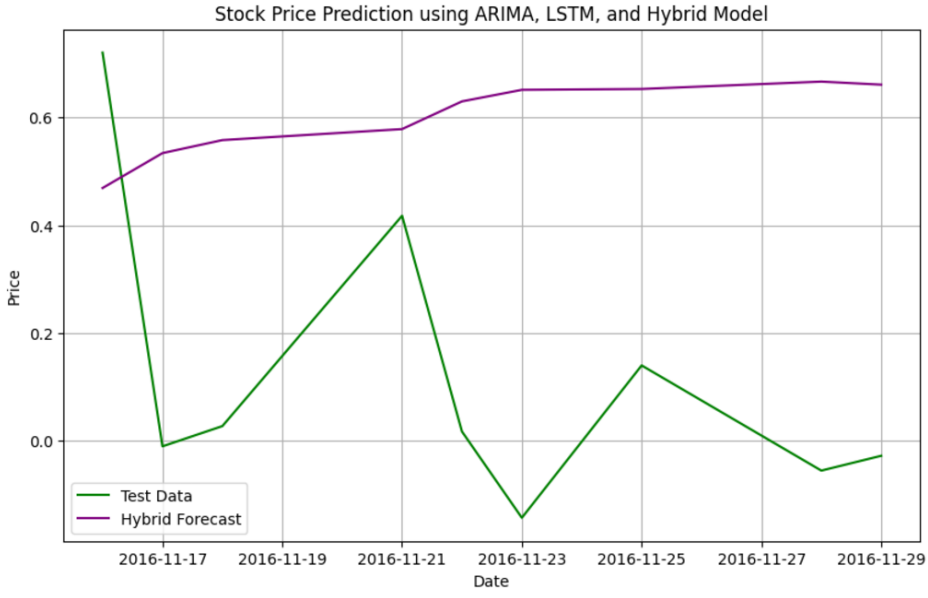
7. Hybrid Model

1. Architecture

- Integration: Combines ARIMA's linear modeling with LSTM's ability to capture non-linear patterns.
- Approach:
  i. ARIMA predictions are generated for the test dataset.
  ii. LSTM predictions are computed for the same dataset.
  iii. Final hybrid prediction: Sum of ARIMA and LSTM predictions (additive approach).

2. Evaluation

- Metrics:
  - Root Mean Squared Error (RMSE): 0.57022
  - Mean Absolute Error (MAE): 0.53481



8. How ARIMA and LSTM Models Complement Each Other

- ARIMA models linear trends and seasonality, excelling in capturing predictable, stationary patterns.
- LSTM models non-linear relationships and long-term dependencies, handling complex and volatile patterns.

Complementing Each Other:

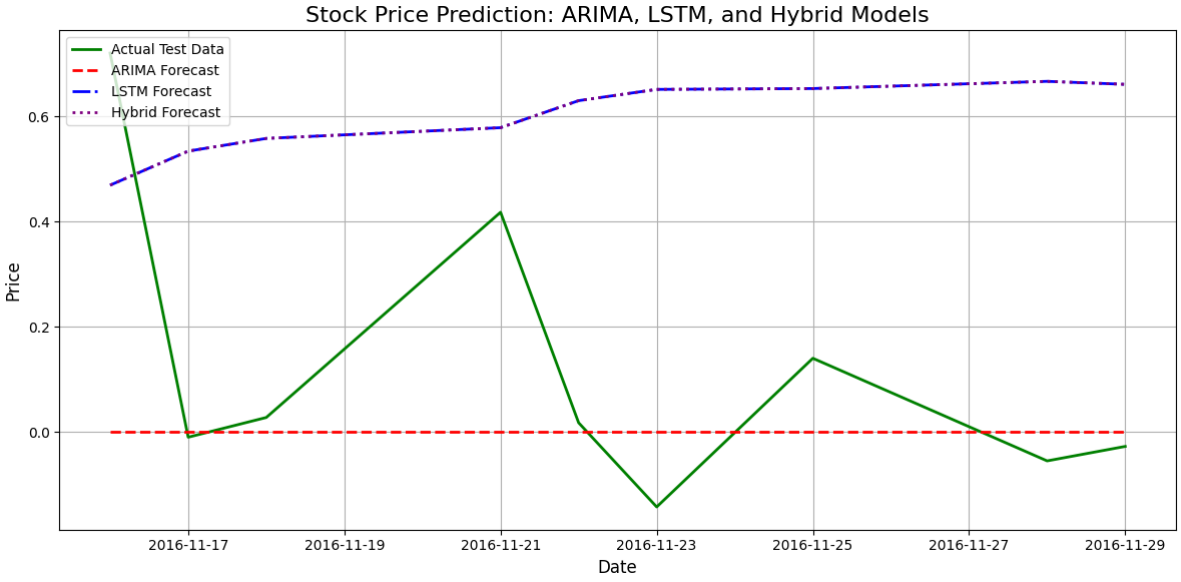
- ARIMA captures the linear components, while LSTM refines the predictions by addressing non-linear trends and residual errors.

Hybrid Model Effect:

- The hybrid model combines ARIMA's linear strengths with LSTM's ability to model complex patterns, leading to more accurate and robust predictions.

9. Conclusion: ARIMA, LSTM, and Hybrid

1. ARIMA Model:
   - Mean Absolute Error (MAE): 0.17306
   - Root Mean Squared Error (RMSE): 0.28627
2. LSTM Model:
   - Mean Absolute Error (MAE): 0.53481
   - Root Mean Squared Error (RMSE): 0.57022
3. Hybrid Model:
   - Mean Absolute Error (MAE): 0.53481
   - Root Mean Squared Error (RMSE): 0.57022



This project successfully forecasts AAPL's stock prices using Auto ARIMA. While the standalone ARIMA model shows reasonable accuracy, implementing LSTM and integrating a hybrid approach is expected to improve predictions.