1. Introduction

The objective of this project was to forecast the future stock values of Apple Inc. (AAPL) using time-series analysis. To achieve this, two distinct modeling approaches were implemented and compared, a traditional statistical model, **ARIMA (Autoregressive Integrated Moving Average)**, and a deep learning model, **LSTM (Long Short-Term Memory)**. The goal was to evaluate their performance and determine which model generalizes better for the volatile and non-linear nature of stock market data.

2. Methodology

2.1. Data

The dataset used for this analysis was the historical daily stock price data for AAPL, sourced from Yahoo Finance. The analysis focused on the 'Close' price from **2015-01-01** to **2023-12-31**.

2.2. Preprocessing

- **ARIMA:** The data was checked for stationarity using the Augmented Dickey-Fuller (ADF) test. As the raw closing prices were non-stationary, first-order differencing was applied to make the series stationary, a prerequisite for the ARIMA model.

- **LSTM:** The data was scaled using a MinMaxScaler to normalize values between 0 and 1. This is a crucial step for deep learning models to ensure stable and efficient training. The data was then transformed into sequences of a specific look-back period (e.g., 60 days) to be used as input features for predicting the next day's price.

2.3. Modeling and Evaluation

Both models were evaluated using a **rolling window (or walk-forward) validation** approach. This method is more robust for time-series data as it simulates a real-world scenario where the model is retrained periodically as new data becomes available.

The following metrics were used to measure forecast accuracy:

- **Root Mean Squared Error (RMSE):** Measures the standard deviation of the prediction errors. It is sensitive to large errors.

- **Mean Absolute Percentage Error (MAPE):** Measures the average percentage error, providing an intuitive measure of forecast accuracy.

3. Results

The performance of the two models on the test data is summarized below.

Model	RMSE	MAPE
ARIMA	15.795903	0.085728
LSTM	5.874788	0.030564

The results clearly indicate that the **LSTM** model provided more accurate predictions, achieving a lower RMSE and MAPE compared to the **ARIMA.**

4. Discussion:

Which Model Generalizes Better and Why?

The **LSTM model demonstrated superior generalization capabilities** for forecasting AAPL stock prices. There are several key reasons for this:

1. **Ability to Capture Non-Linearity:** Stock market data is inherently non-linear and complex, influenced by market sentiment, news, and complex economic factors. ARIMA is a linear model that assumes a linear relationship between past observations. It struggles to capture the complex, non-linear patterns present in financial time series. The LSTM, being a deep neural network, excels at learning and modeling these intricate, non-linear dependencies.

2. **Long-Term Memory:** LSTMs are specifically designed to overcome the vanishing gradient problem in standard RNNs, allowing them to learn long-term dependencies in sequential data. The "memory cells" in an LSTM can retain information over long periods. This is a significant advantage in financial markets, where events from months ago can still influence current trends. ARIMA's memory is limited by its p and qparameters, making it more suited for shorter-term, more stable patterns.

3. **Adaptability to Volatility:** The stock market is characterized by periods of high and low volatility. The LSTM model, by learning from sequential data windows, can better adapt to these changing dynamics. The linear assumptions of ARIMA are often violated during sudden market shifts, leading to poorer performance.

While the ARIMA model serves as an excellent statistical baseline, its limitations prevent it from effectively generalizing to the dynamic and often unpredictable nature of stock price movements. The LSTM's architecture is fundamentally better suited for this type of data, allowing it to produce more robust and accurate forecasts on unseen data.

5. Conclusion and Recommendation

Based on the quantitative results (lower RMSE and MAPE) and the qualitative analysis, the **LSTM model is the recommended choice** for forecasting AAPL stock prices. Its inherent ability to model complex, non-linear relationships and long-term dependencies makes it a more powerful and generalizable tool for this task compared to the traditional ARIMA model.