Prepared by group 5

STOCK STRATEGY

BRIANNA, EMMA, JULIANA & SIDNEY

April 14th, 2025

STA 4365

AGENDA

- Introduction
- Methodology
- Data Source
- Feature Engineering
- LSTM
- ARIMAX
- CNN
- XGBoost
- Model Analysis
- Conclusion

INTRODUCTION

Each member of our group implemented a different model to forecast Q1 closing prices for AAPL— including **LSTM, XGBoost, CNN, and ARIMAX**. To ensure a fair comparison, we first evaluated all models using a static approach: training on historical data through 2024 and generating predictions for each trading day in Q1 2025. After comparing performance metrics of all 3 models, we found that the XGBoost model performed best overall. Because daily retraining is computationally intensive, we chose to apply **daily retraining only to the best-performing model — XGBoost** — to simulate how a real trading system might incrementally adapt.

METHODOLOGY : APPLE STOCK (AAPL)

FEATURE ENGINEERING:

- Each team member selected a unique combination of engineered features—such as open, high, low, close, volume, moving averages (5-day and 10-day), percent change, and RSI—to build their model inputs.

MODEL DEVELOPMENT:

- Using these features, each member trained a separate model to forecast the closing price of AAPL for every trading day in Q1 2025.
- All models used only information that would be available at the end of each trading day to simulate real-world post-market close forecasting conditions.

PERFORMANCE EVALUATION:

- We evaluated each static model's predictions using key performance metrics: RMSE, MAE, R^2, and Directional Accuracy.
- This allowed for a fair comparison of both error magnitude and trend prediction effectiveness.

MODEL SELECTION FOR LIVE FORECASTING:

- The model with the best overall performance will be used going forward to predict the next trading day’s closing price using the most recent available features—for example, using Monday’s market data to predict Tuesday’s close.

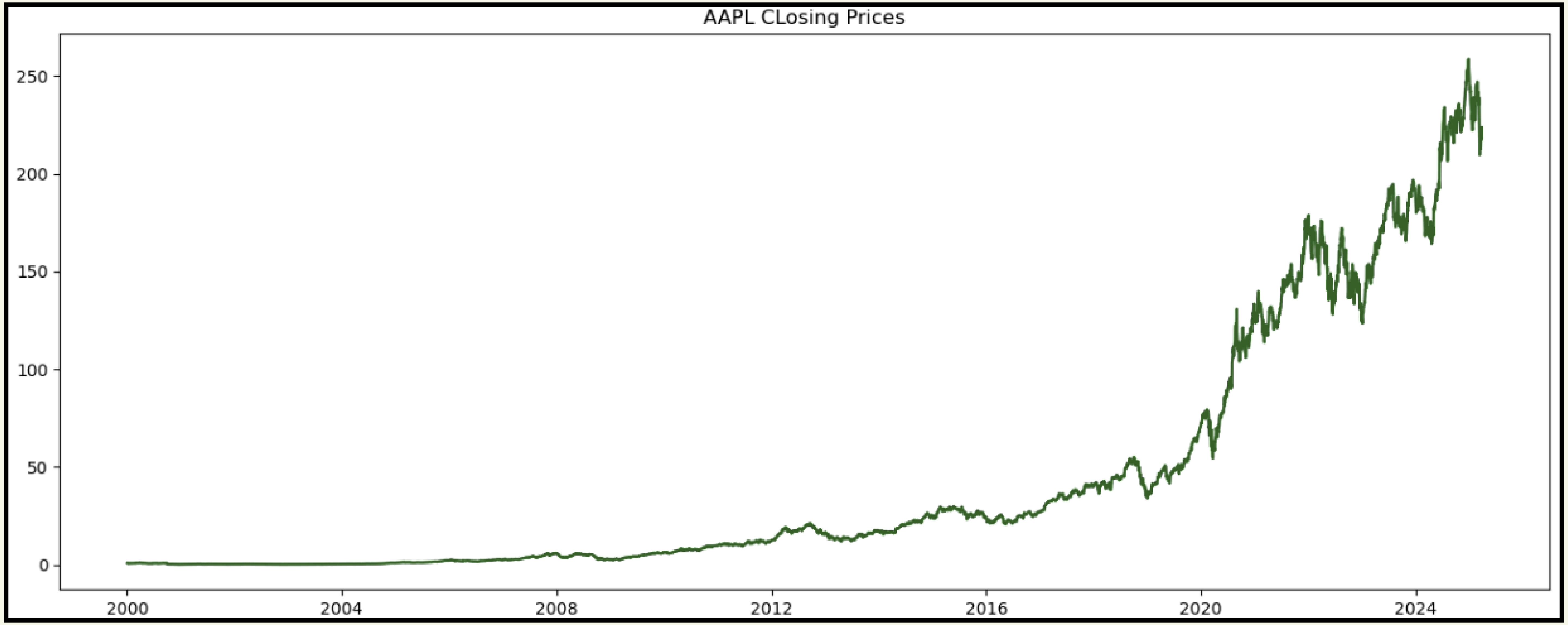
DATA SOURCE

YFINANCE

Open source library containing data from Yahoo Finance.

Date	Close	High	Low	Open	Volume
3/31/2020	61.71085739	63.70082497	61.15512397	62.02877007	197002000
4/1/2020	58.46380234	60.35912485	58.03183399	59.82037714	176218400
4/2/2020	59.43937683	59.49276652	57.49066439	58.3254809	165934000
4/3/2020	58.58513641	59.62622766	57.99299908	58.92246006	129880000

AAPL over the years



- **Historically upward trend**
- **Increased volatility after COVID:** Lockdowns gave people more free time, many used stimulus checks to start trading, leading to a major shift in retail investor behavior — clearly reflected in AAPL stock.
- **Market Crash:** Buying the dips became common, with many inexperienced traders jumping in (especially on zero commission platforms like Robinhood), which contributed to sharper and more frequent price swings.

Our Models

Let the predictions begin

FEATURE ENGINEERING

Close_lag1	Closing price 1 day ago
Open_lag1	Opening price 1 day ago
Low_lag1	Lowest price 1 day ago
High_lag1	Highest price 1 day ago
MA_lagx	Moving average of closing price across x days
Volume_lag1	Trading volume 1 day ago
Return_1d	Daily price percentage change
RSI_14	Speed and magnitude of recent price changes over the last 14 trading days

Long Short-Term Memory (LSTM) RNN

Features used: Previous day's Closing Price, 5-day MA, 10-day MA, Daily Percentage Return, 14-day Relative Strength Index (RSI)

Architecture (built with Keras):

Training: 2010-01-02 : 2024-12-31

Epochs: 10

Uses features from previous 60 days

Layer (type)	Output Shape
lstm (LSTM)	(None, 60, 128)
lstm_1 (LSTM)	(None, 64)
dense (Dense)	(None, 25)
dense_1 (Dense)	(None, 1)

1st LSTM Layer has 128 neurons

2nd LSTM Layer has 64 neurons

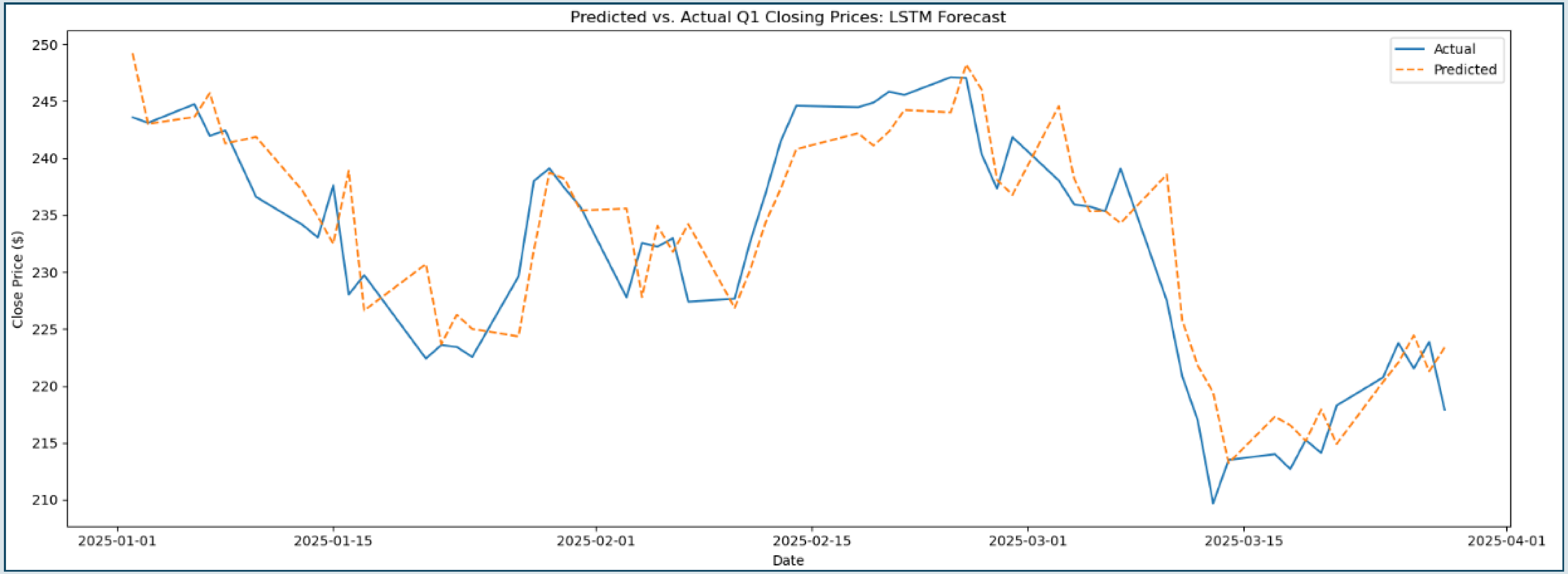
1st Dense Layer has 25 neurons

One output (predicted close)

Something cool I learned while tuning hyperparameters was how important it is to understand the **context** and nature of your data. Stock prices are volatile, and using a batch size of 1 allowed the model to adapt to daily fluctuations by updating weights more frequently. This made the model more sensitive to short-term trends that might have been smoothed out with larger batch sizes.

Long Short-Term Memory (LSTM) RNN

Q1 Predictions vs. Actual Close



RMSE: 4.40

MAE: 3.44

R-Squared: 0.82

Directional Accuracy: 50.12%

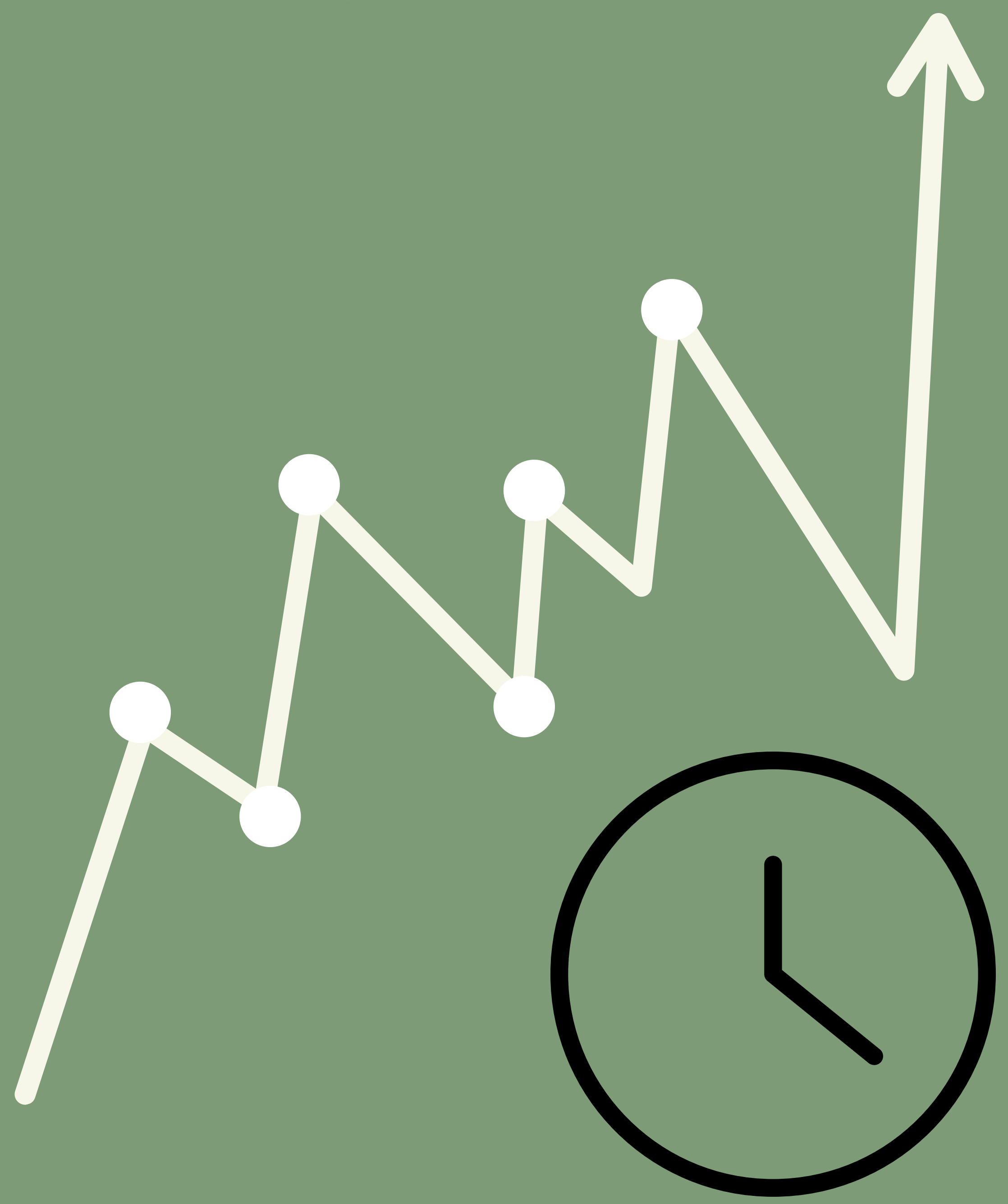
The LSTM model captures AAPL's overall trend well but reacts slowly to sharp price changes. It's accurate during stable periods but could be improved to better handle volatility.

ARIMAX

Autoregressive Integrated Moving Average with Exogenous Features

- Similar to an *ARIMA (p, d, q)* model but incorporates *exogenous variables*
- Captures both: *time* dependent patterns & influence from *external factors*

Parameter	Name	Usage
p	Autoregressive (AR)	How many past values the model uses to predict the next one
d	Differencing (I)	How many times the data is differenced to make it stationary
q	Moving Average (MA)	How many past forecast errors the model uses for correction



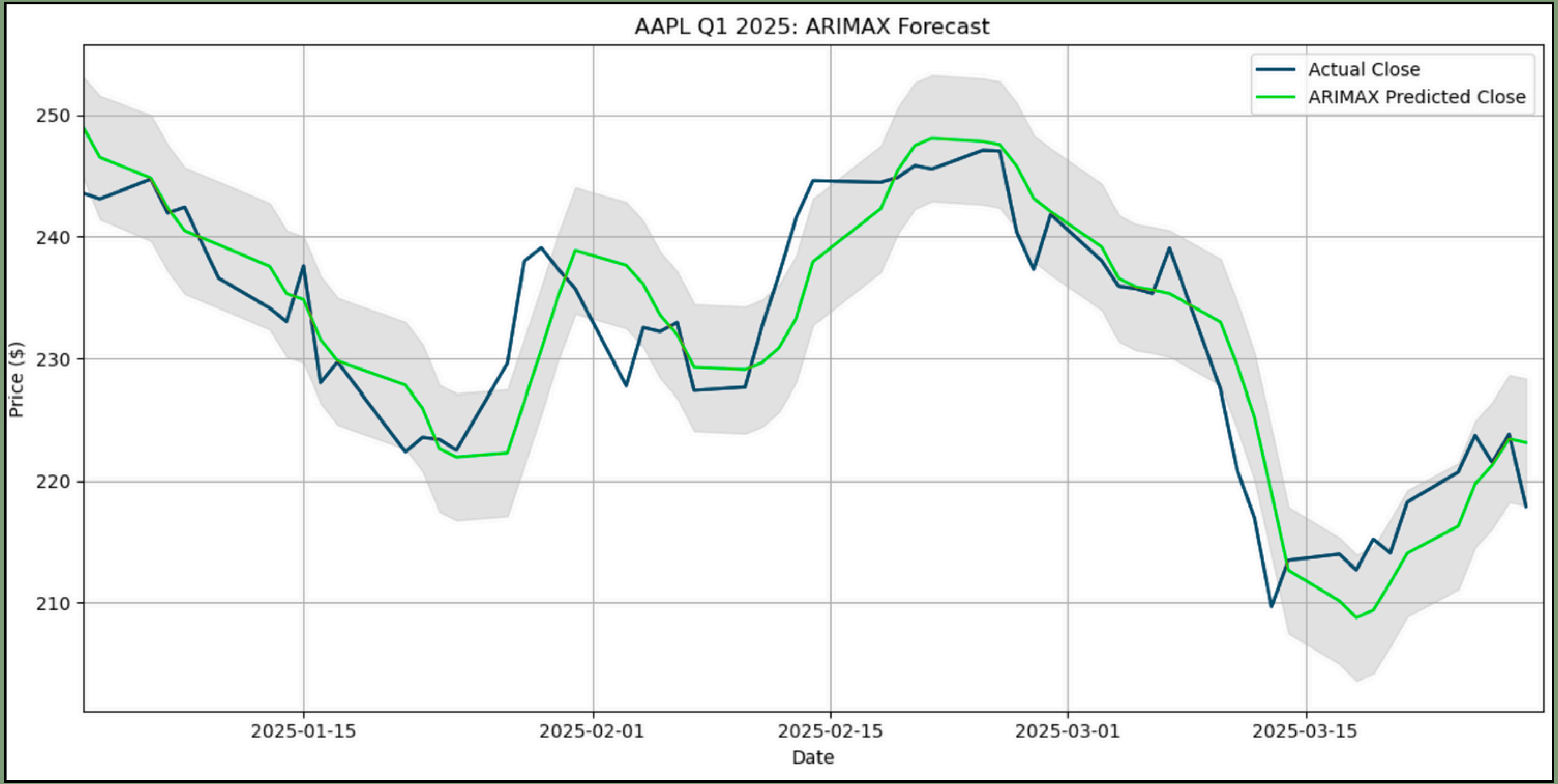
ARIMAX

Features used: Volume_lag1, (High_lag1 - Close_lag1), 5-day MA, 10-day MA → Feature scaling

Training Period: 1/01/2020 - 12/31/2024

Testing Period: Q1 2025 (1/01/2025 - 3/28/2025)

Parameters: (p, d, q) → (2, 1, 1)



RMSE: 4.47

MAE: 3.45

R-Squared: 0.81

Directional Accuracy: 67.24%

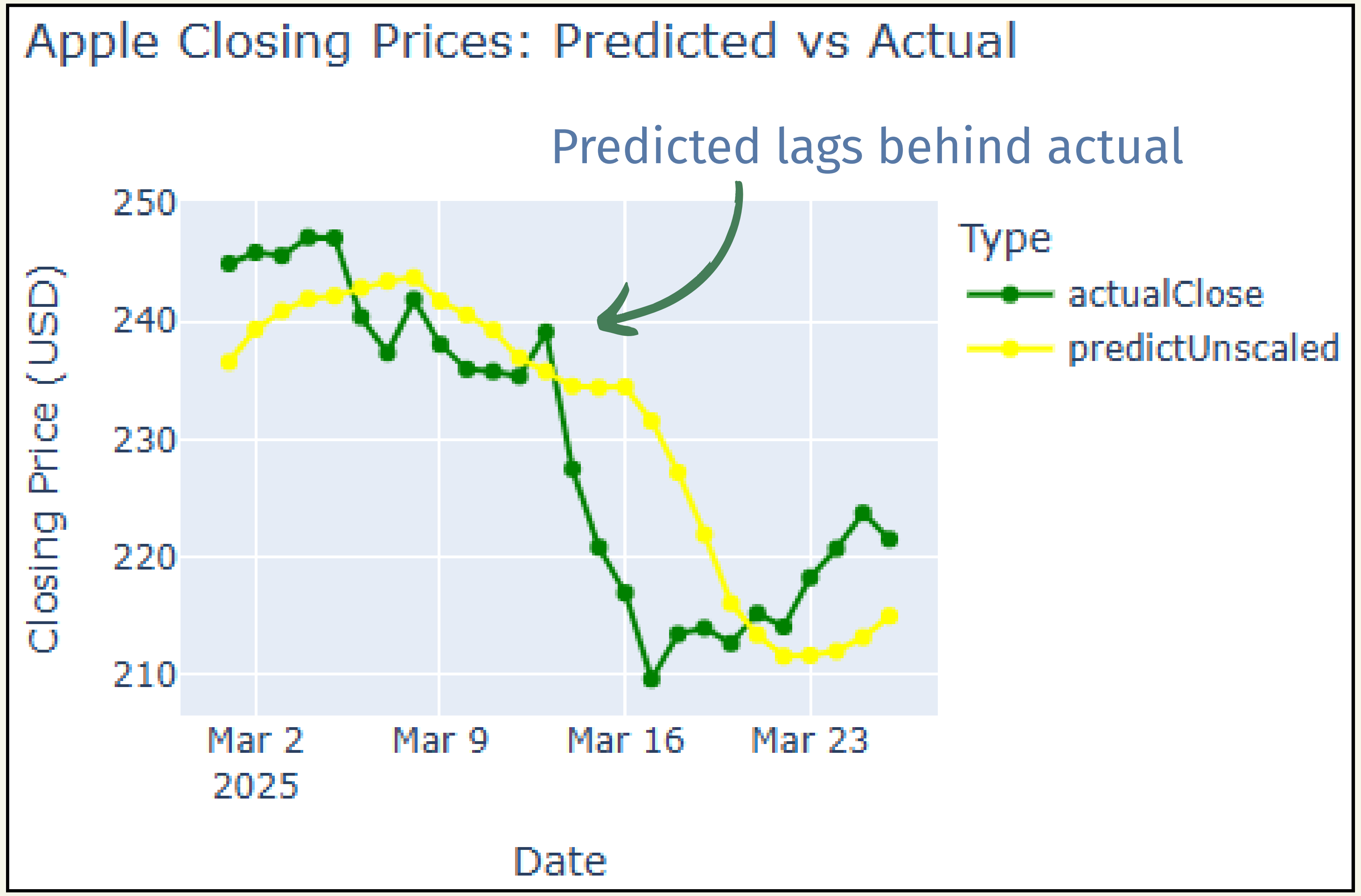
CNN

While commonly used for images, the sequential nature of CNN means it can be used to find patterns in time-series data.

Time Period: 3/31/2020 - 3/26/2025

Curious about using only close to predict future close.

- **Stock prices fluctuate a lot throughout the day**



RMSE: 8.43

MAE: 6.85

R-Squared: 0.55

Directional Accuracy: 57.69%

Only using close helps with generalization and prevent data leakage that could occur with variables like open, high, and low.

CNNs are good at finding patterns, but for something with as much volatility as stocks, the CNN can't entirely fit to this.

XGBOOST

Features used: Lagged Closing Price for days 1-5, 5-day MA, 10-day MA, Yesterday's Return, 14-day Relative Strength Index (RSI)

Training Period: Static - Stopped training Dec 2024

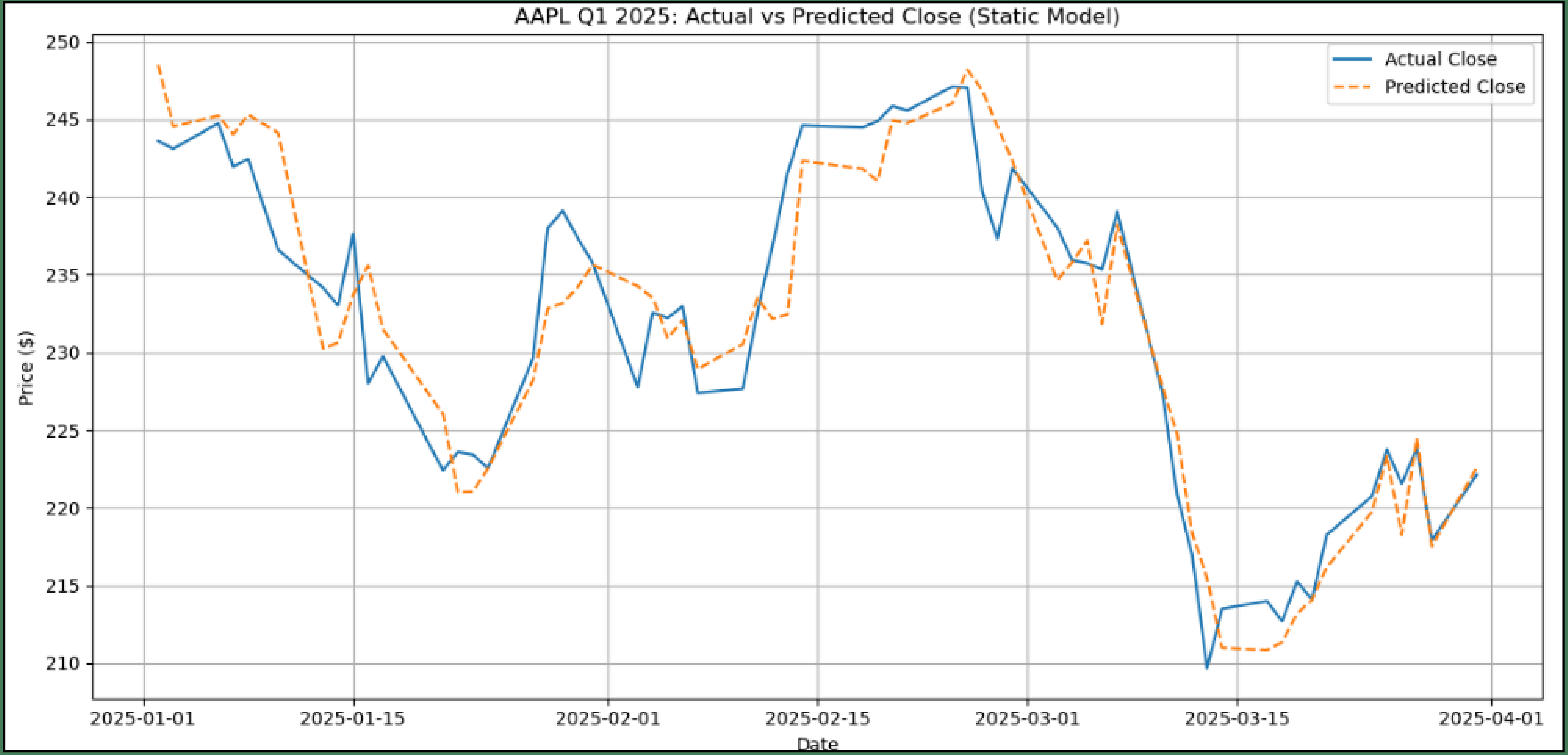
Static Model

RMSE: 3.41

MAE: 2.62

R-Squared: 0.89

Directional Accuracy: 67.80%



MODEL ANALYSIS

buy or sell?

hmm...

Model	RMSE	MAE	R^2	Directional Accuracy
XGBoost (winner!)	3.41	2.62	0.89	67.80
ARIMAX	4.47	3.45	0.81	67.24%
LSTM	4.20	3.39	0.83	50.12%
CNN	8.43	6.85	0.55	57.69%

XGBOOST

Features used: Lagged Closing Price for days 1-5, 5-day MA, 10-day MA, Yesterday's Return, 14-day Relative Strength Index (RSI)

Training Period: Continuous - retraining model daily

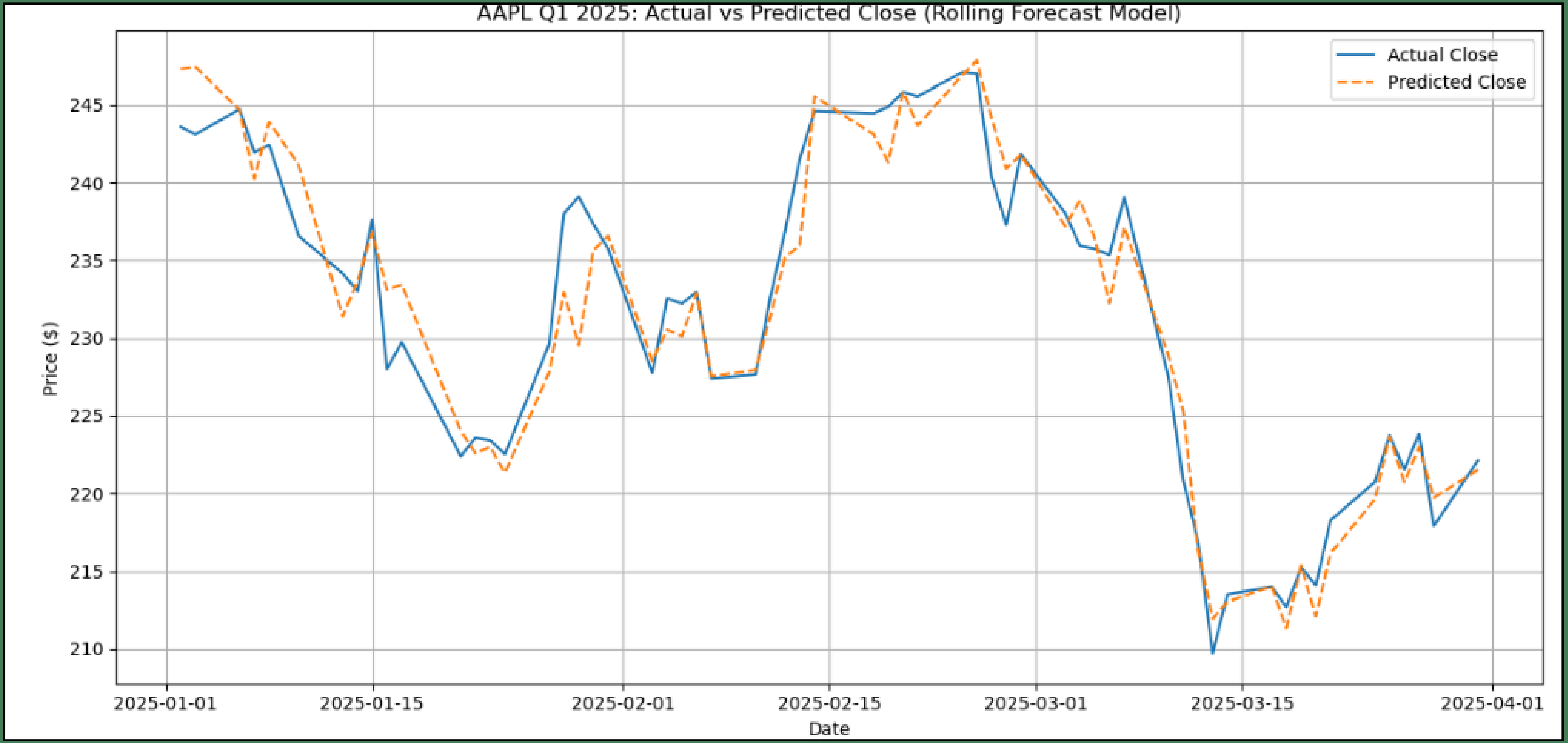
Rolling Forecast

RMSE: 2.57

MAE: 1.86

R-Squared: 0.94

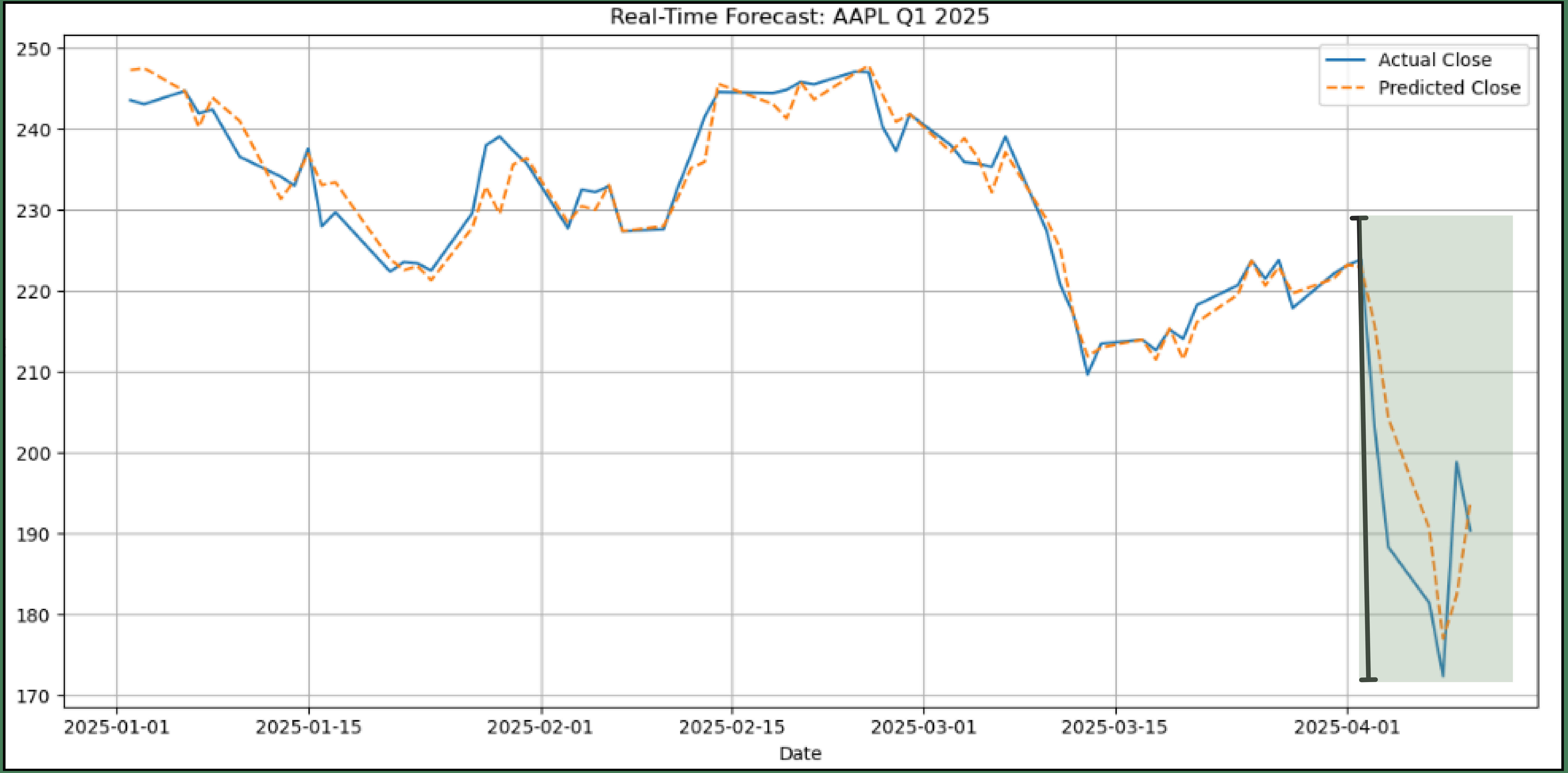
Directional Accuracy: 81.36%



XGBOOST

Features used: Lagged Closing Price for days 1-5, 5-day MA, 10-day MA, Yesterday's Return, 14-day Relative Strength Index (RSI)

4 day Trial

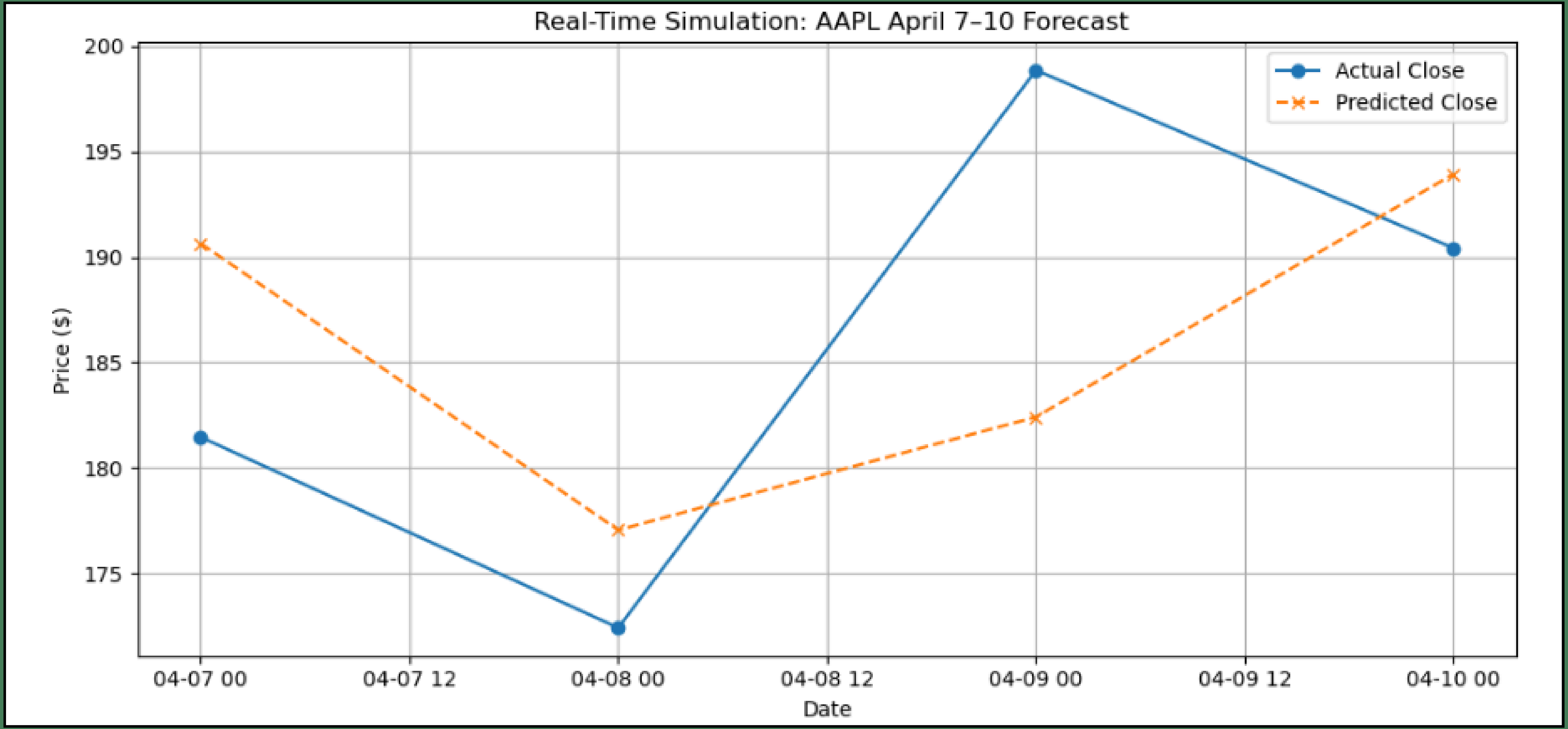


XGBOOST

Features used: Lagged Closing Price for days 1-5, 5-day MA, 10-day MA, Yesterday's Return, 14-day Relative Strength Index (RSI)

```
Forecast for April 7-10, 2025
            Predicted_Close  Actual_Close
Date
2025-04-07       190.642548    181.460007
2025-04-08       177.055313    172.419998
2025-04-09       182.415375    198.850006
2025-04-10       193.930038    190.419998

 Evaluation Metrics (4-day real forecast):
RMSE: 9.85
MAE: 8.44
R² Score: 0.0033
Directional Accuracy (April 7-10): 66.67%
```



CONCLUSION

Objective: Predict short-term Apple stock prices.

Data Source: Yahoo Finance (2020–2025)

Best Model: XGBoost with Rolling Forecasting

Performance: RMSE: 2.57 , R^2 Score: 0.94 , and Directional Accuracy: 31.36%

Use Case: Short-term prediction engine and sell recommendation signal

Next Steps: Add more helpful data like **market news**, trading volume, or **economic trends** to improve predictions

Next 4 days Prediction | Rolling XGBOOST

```
2025-04-14: $192.73
2025-04-15: $183.32
2025-04-16: $186.13
2025-04-17: $192.88
```

THANK YOU

DEFINITIONS

`Close_lag1` to `Close_lag5`	Past closing prices	"What the stock closed at 1–5 days ago"	If today is April 6: `Close_lag1 = April 5 close`
`MA_5` (5-day Moving Average)	Short-term price trend	"Average closing price over the last 5 days"	$MA_5 = (Close_{t-1} + Close_{t-2} + ... + Close_{t-5})/5$
`Return_1d`	1-day return	"How much the price changed from yesterday to today, in %"	$\text{Return\_1d} = \frac{Close_t - Close_{t-1}}{Close_{t-1}}$ Example: If yesterday's price was \$100 and today is \$105: $\frac{105-100}{100} = 0.05 = 5\%$
`RSI_14` (Relative Strength Index)	Momentum indicator	"Shows if the stock might be overbought or oversold"	Uses 14 days of price changes: $RSI = 100 - \frac{100}{1+RS}$ where RS = Avg Gain / Avg Loss

SOURCES

- https://www.forbes.com/councils/forbesbusinesscouncil/2024/01/30/state-of-retail-trading-the-evolving-retail-trading-landscape/
- https://www.techradar.com/news/coronavirus-stimulus-checks-used-to-buy-stocks
- https://yfinance-python.org/reference/index.html