TCS Stock Data

Live and Latest (With yfinance)

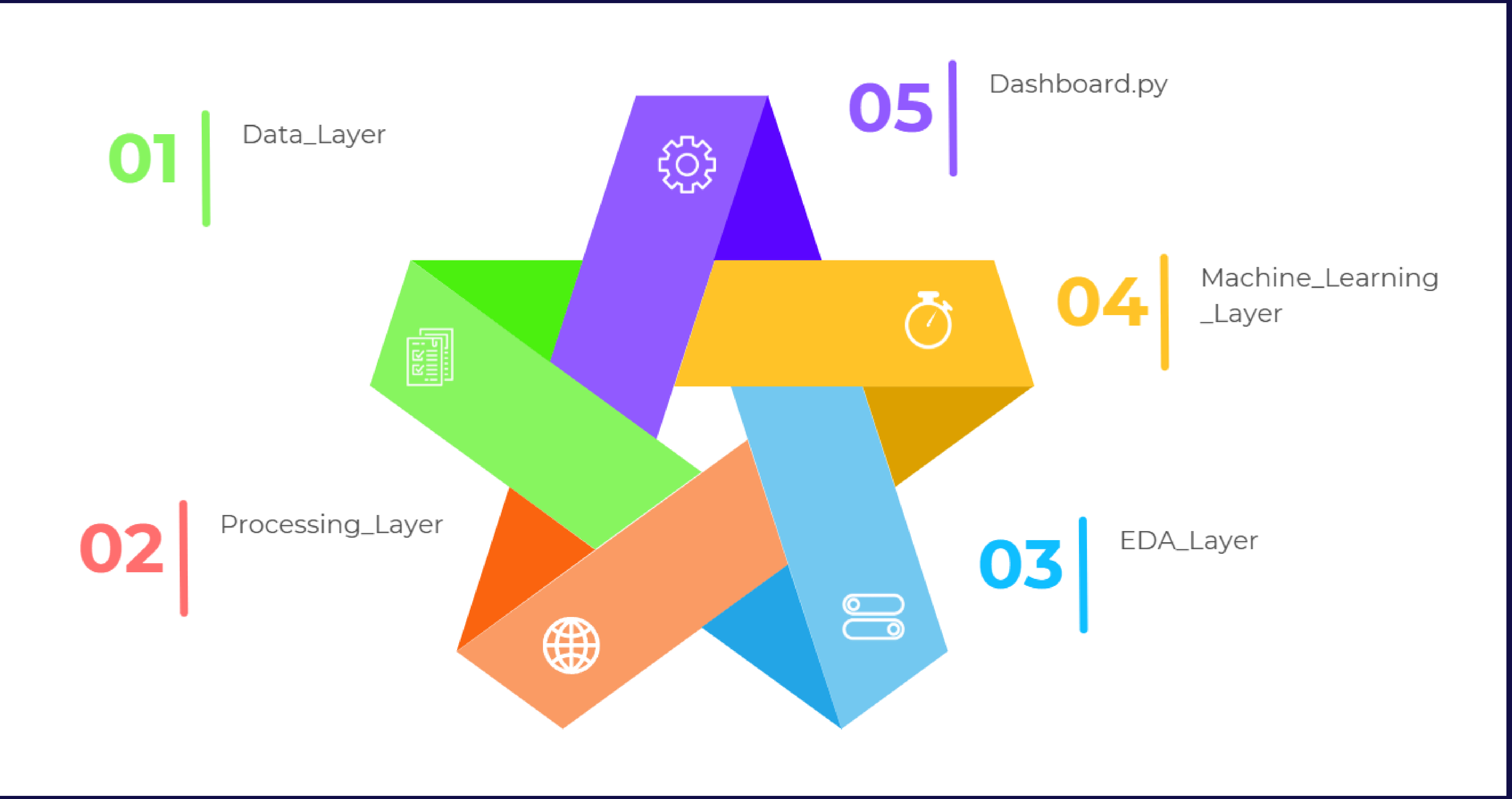
OIKANTIK BASU
Data Science Intern at Unified Mentor

MAY, 2025 @ 09:00 A.M.

OBJECTIVE

The objective of this project is to analyze and forecast TCS stock prices by leveraging historical market data, technical indicators, and machine learning models. Through comprehensive feature engineering and exploratory analysis, the project aims to capture short- and long-term market behavior, assess volatility and momentum, and evaluate the predictive accuracy of models like Random Forest, XGBoost, and LSTM. The ultimate goal is to build a reliable and interpretable system that aids in short-term price prediction while highlighting the limitations of model performance over longer durations.

System Architecture Star Diagram



Dataset Overview

TCS_1D_1MIN.CSV

A high-frequency dataset containing 1-minute interval price data for TCS over a single trading day. Useful for intraday trend detection, micro-level trading pattern analysis, and real-time price movement modeling.

TCS_5D_30MIN.CSV

This dataset holds 30-minute interval OHLCV data across 5 recent trading days. It helps in short-term trend tracking, momentum-based strategy analysis, and session-level pattern identification.

TCS_1M_1D.CSV

A medium-range dataset covering daily trading information for TCS stock over the past month. Suitable for short-term forecasting models, technical analysis, and backtesting quick strategies.

TCS_6M_1D.CSV

Daily stock data for TCS over 6 months. Ideal for analyzing quarterly trends, seasonality, and evaluating recent performance before earnings cycles or news impact.

TCS_1Y_1D.CSV

This file contains a year's worth of daily OHLCV data. It enables annual performance evaluation, technical indicator computation (e.g., RSI, MACD), and trend-based ML model training.

TCS_5Y_7D.CSV

A long-term view of TCS stock with weekly aggregated data spanning 5 years. Best for macro-trend analysis, investment horizon modeling, and visual storytelling over years.

NIFTY50.CSV

Index-level daily price data of NIFTY 50, which serves as a market benchmark. Useful for comparative analysis with TCS, sector correlation studies, and market direction context.

USDINR.CSV

Contains historical USD to INR exchange rate data. Crucial for studying the influence of forex rates on export-driven companies like TCS, especially in global revenue exposure models.

Feature Engineering

PRICE

The average of high and low prices for the day. Used to smooth out sharp fluctuations in OHLC data and is often the base for calculating indicators like Bollinger Bands and Moving Averages.

MA20 & MA50

Moving Averages calculated over 20 and 50 days, respectively. These help identify short- and mid-term trends. Crossovers between them often signal potential buy or sell points.

RSI

A momentum oscillator ranging from 0 to 100. Values above 70 suggest an overbought market, while values below 30 indicate oversold conditions, aiding in reversal detection.

MACD

Shows the relationship between two EMAs (typically 12- and 26-day). It indicates trend strength and direction, and is commonly used with the Signal Line to trigger trades.

SIGNAL LINE

The 9-day EMA of MACD. Often plotted on top of MACD to show crossovers that may indicate bullish or bearish shifts in trend momentum.

BOLLINGER BANDS (HIGH & LOW)

Calculated as a 20-day moving average plus/minus 2 standard deviations. These bands show volatility and potential price reversals when the close price touches or breaches the bands.

LAG FEATURES (LAG_1 TO LAG_7)

Closing prices shifted backward by 1 to 7 days. These are essential for time-series models like LSTM, ARIMA, or Random Forest to learn from historical price dependencies.

VOLATILITY_5D

The rolling 5-day standard deviation of returns. It captures short-term market turbulence and helps quantify risk for model training or portfolio management.

PRICE_CHANGE

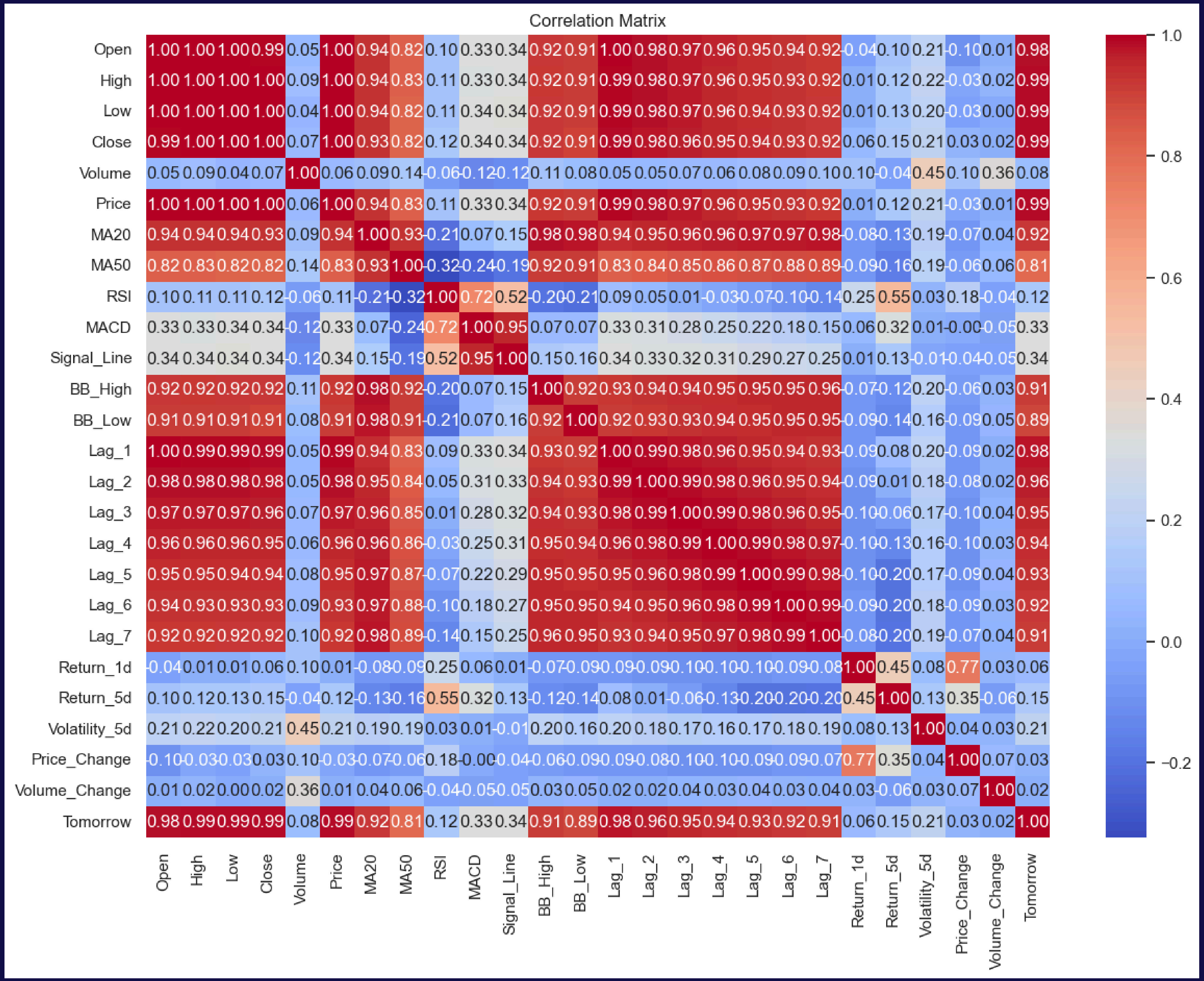
Represents the net change in price from the previous day. It gives a raw directional indicator of market movement before normalization or scaling.

VOLUME_CHANGE

Shows the difference in traded volume compared to the previous day. Helps understand whether price moves are backed by significant market interest or not.

Exploratory Data Analysis

The exploratory data analysis (EDA) of TCS stock reveals rich patterns in price behavior, market momentum, and volatility. The correlation matrix highlights that core features like Open, High, Low, Close, and Price are strongly interrelated and well-aligned with the target variable (Tomorrow), while indicators such as RSI, MACD, Volume_Change, and Volatility_5d show weaker direct correlations. Moving averages (MA20 & MA50) clearly track trends, with MA crossovers acting as reliable buy/sell signals. RSI and MACD add momentum insights RSI's 68 overbought and 70 oversold days suggest high volatility, while 38 MACD crossovers signal frequent trend shifts. Bollinger Bands further emphasize periods of high and low volatility, often aligning with price reversals. A look at rolling volatility and monthly close prices uncovers seasonal cycles and instability spikes, especially post-Jan 2023. Autocorrelation plots (ACF & PACF) justify the use of lag-based features in modeling. Volume analysis reveals erratic behavior, with limited correlation to price but relevance during spikes. Daily return distributions resemble a normal curve with fat tails, affirming the need for robust models. Lastly, cumulative returns indicate a profitable trend till early 2023, followed by a downturn signaling the importance of dynamic models that adjust to regime shifts.



Correlation Matrix

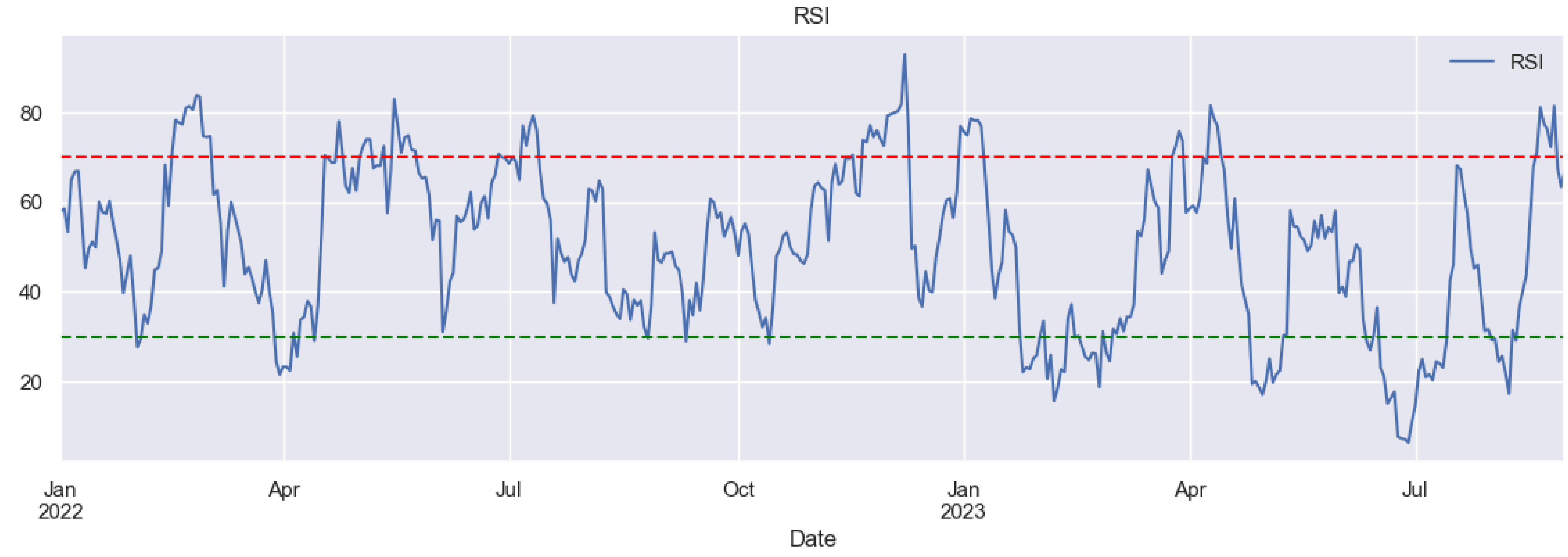
Correlation Matrix

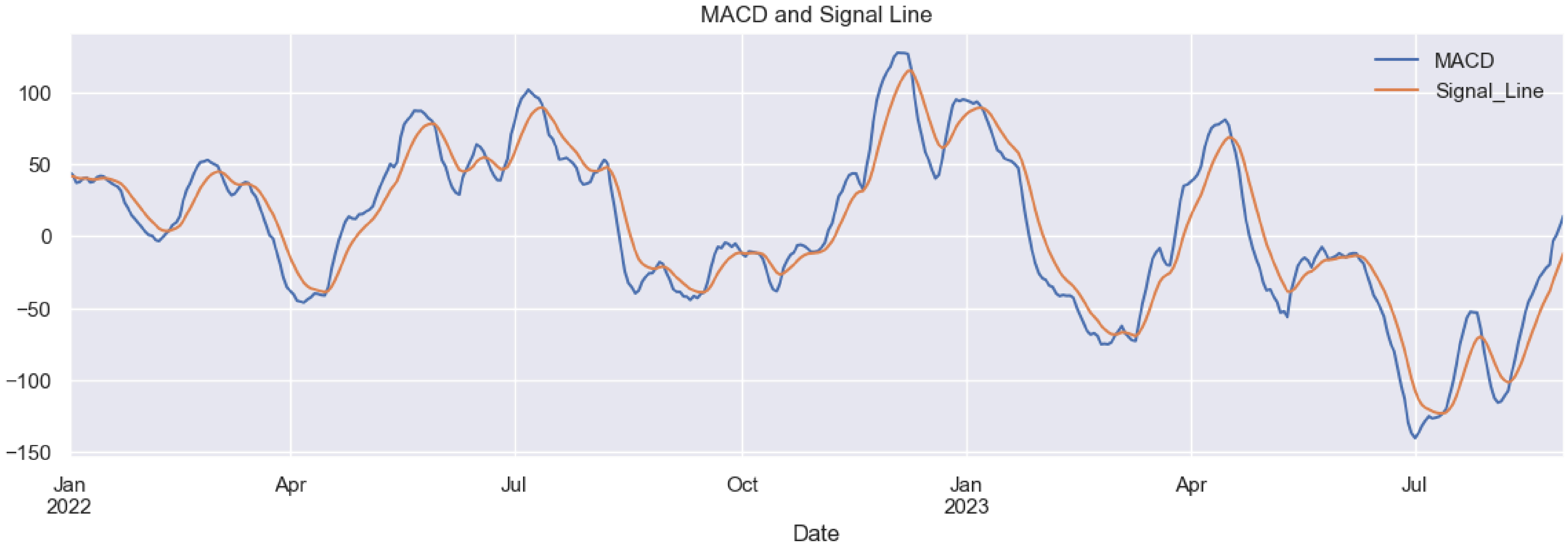
- Highly correlated features: Open, High, Low, Close, Price, and moving averages (MA20, MA50) all show strong positive correlation with each other and the Tomorrow value.
- Low correlation features: RSI, MACD, Signal_Line, Volume_Change, and Volatility_5d show weak correlation with the target (Tomorrow), suggesting limited direct predictive value.

Price with Moving Averages (MA20 & MA50)

MA20 & MA50

- Trend tracking: Price follows a cyclical pattern with visible uptrends and downtrends. MA crossovers (especially when MA20 crosses MA50) can signal potential buy/sell points.
- Bearish signal: Prolonged periods where price stays below MA50 are indicative of downtrends (e.g., after Jan 2023).





RSI & MACD

RSI & MACD

RSI (Relative Strength Index):

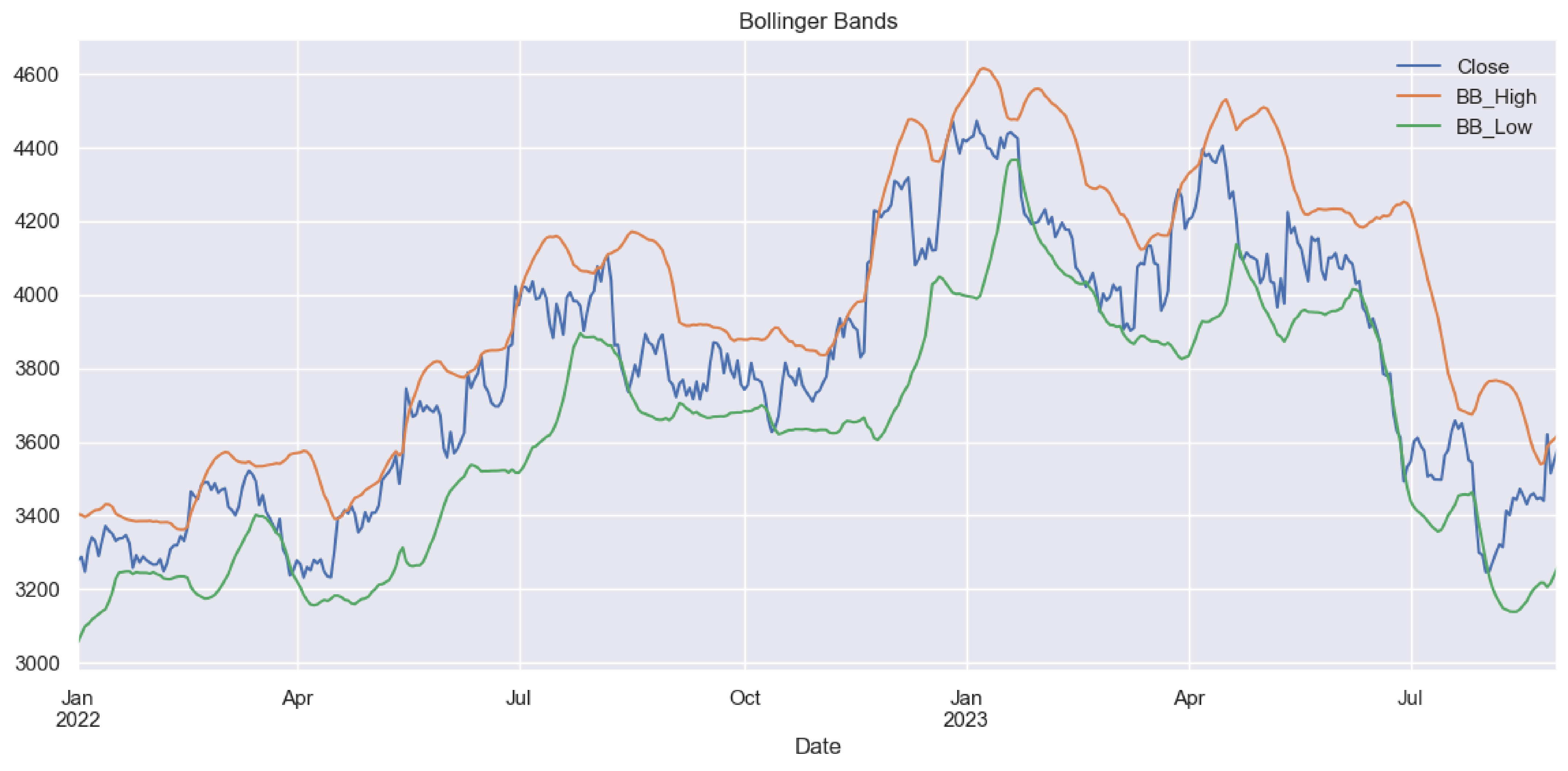
- Overbought zone: Above 70 (seen multiple times, e.g., early 2023)
- Oversold zone: Below 30 (notable around June–July 2023)
- Useful for short-term trend reversals.

MACD & Signal Line:

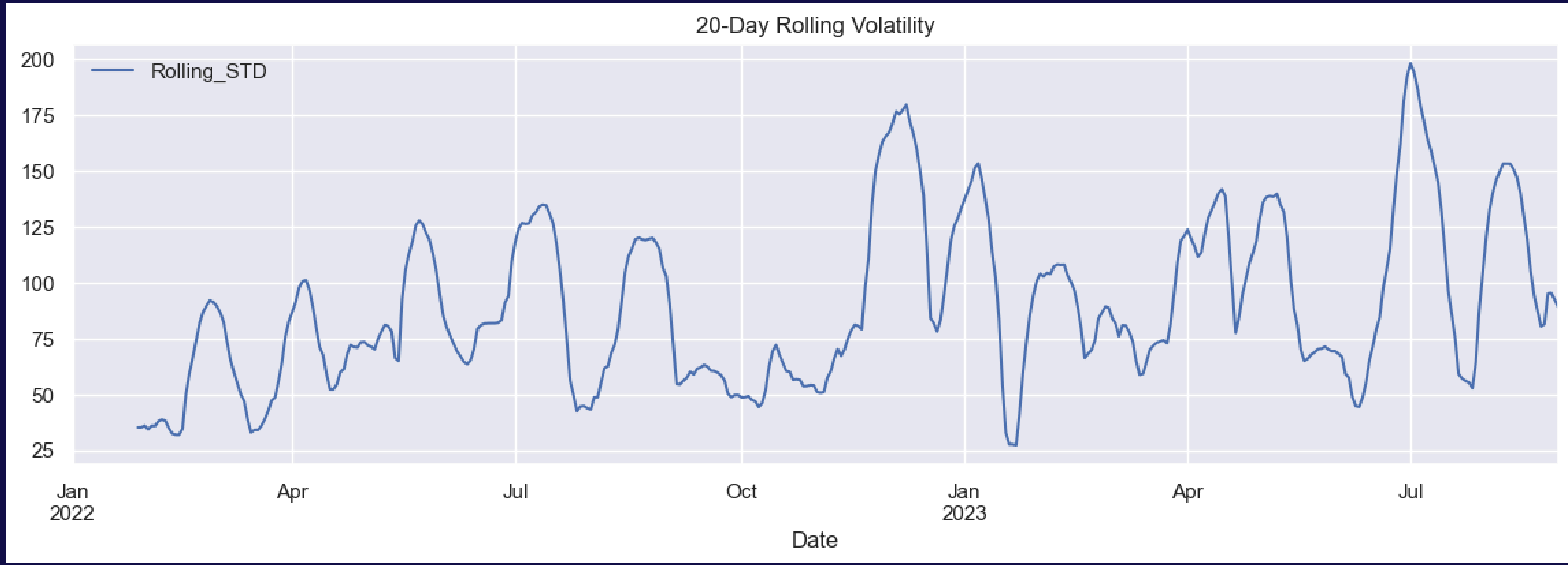
- Convergence/divergence patterns are evident.
- Bullish crossover: MACD > Signal (early signs of trend reversal).
- Bearish crossover: MACD < Signal.

Bollinger Bands

- Volatility indicator: Wider bands suggest high volatility, while tighter bands suggest consolidation.
- Price touching upper/lower bands can signal potential reversal or breakout points.



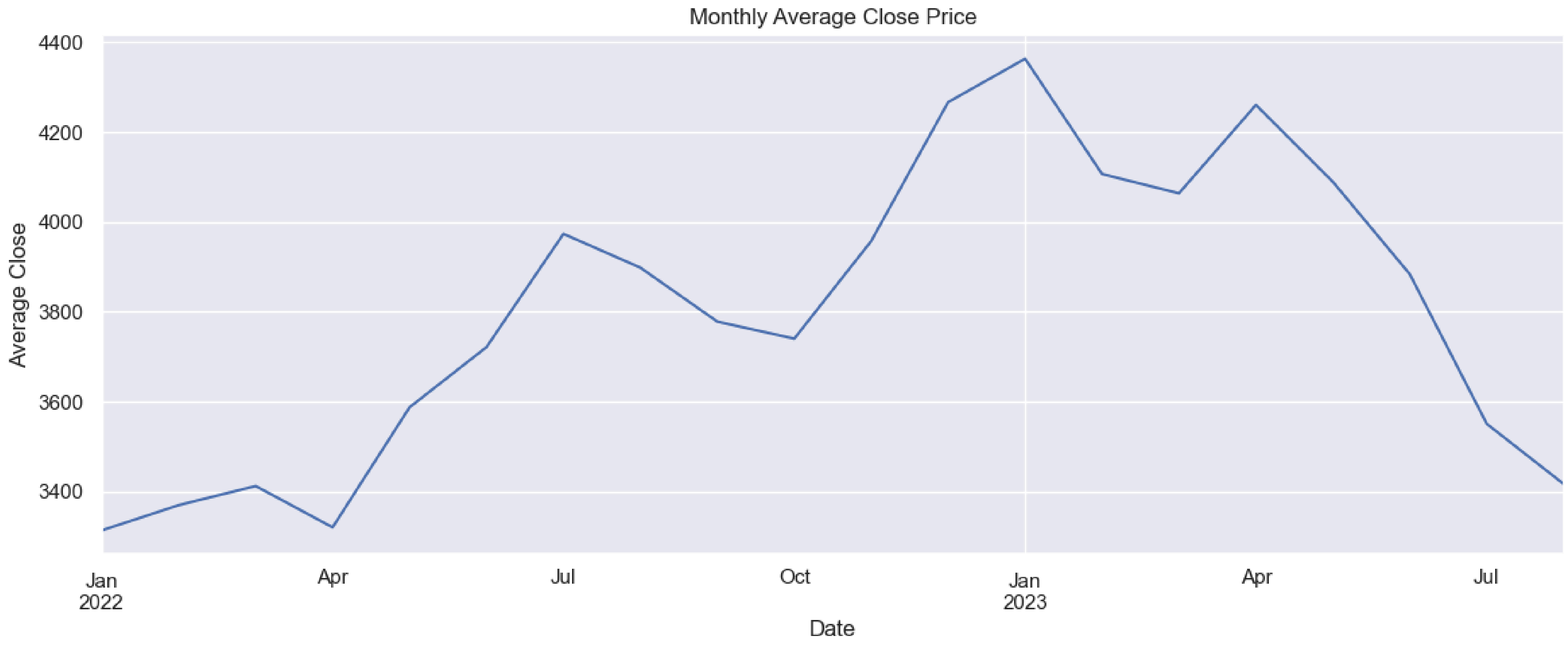
Bollinger Bands



Volatility

Volatility

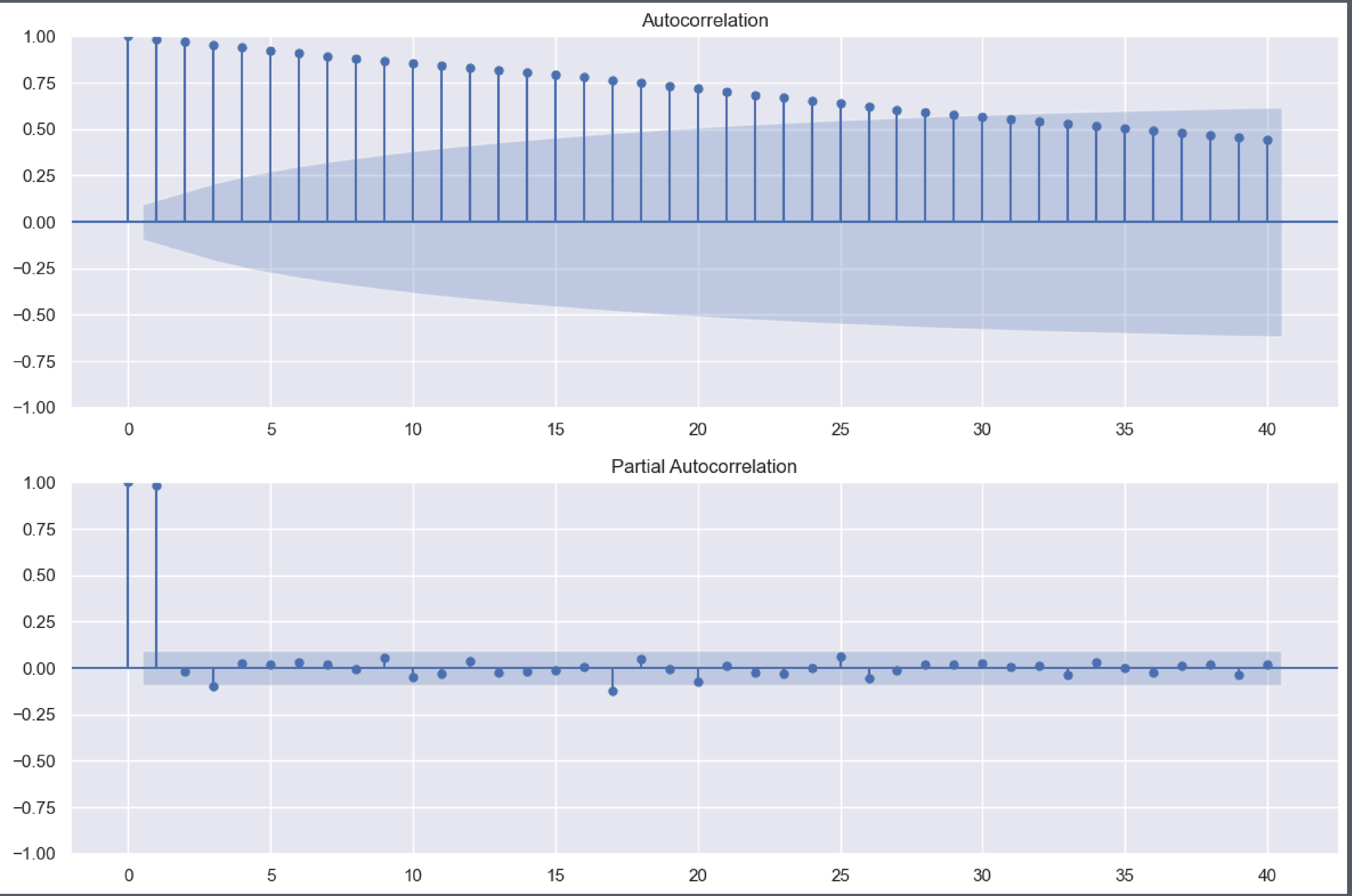
- Spikes in rolling volatility (especially Dec 2022 & mid-2023) indicate unstable periods—ideal for model focus on risk-adjusted return predictions.



Monthly Close Price

Monthly Close Price

- Clear seasonality: Price surged from mid-2022 to early 2023, followed by a consistent downtrend.
- Trend insights: Model can incorporate seasonality or rolling window-based trends.

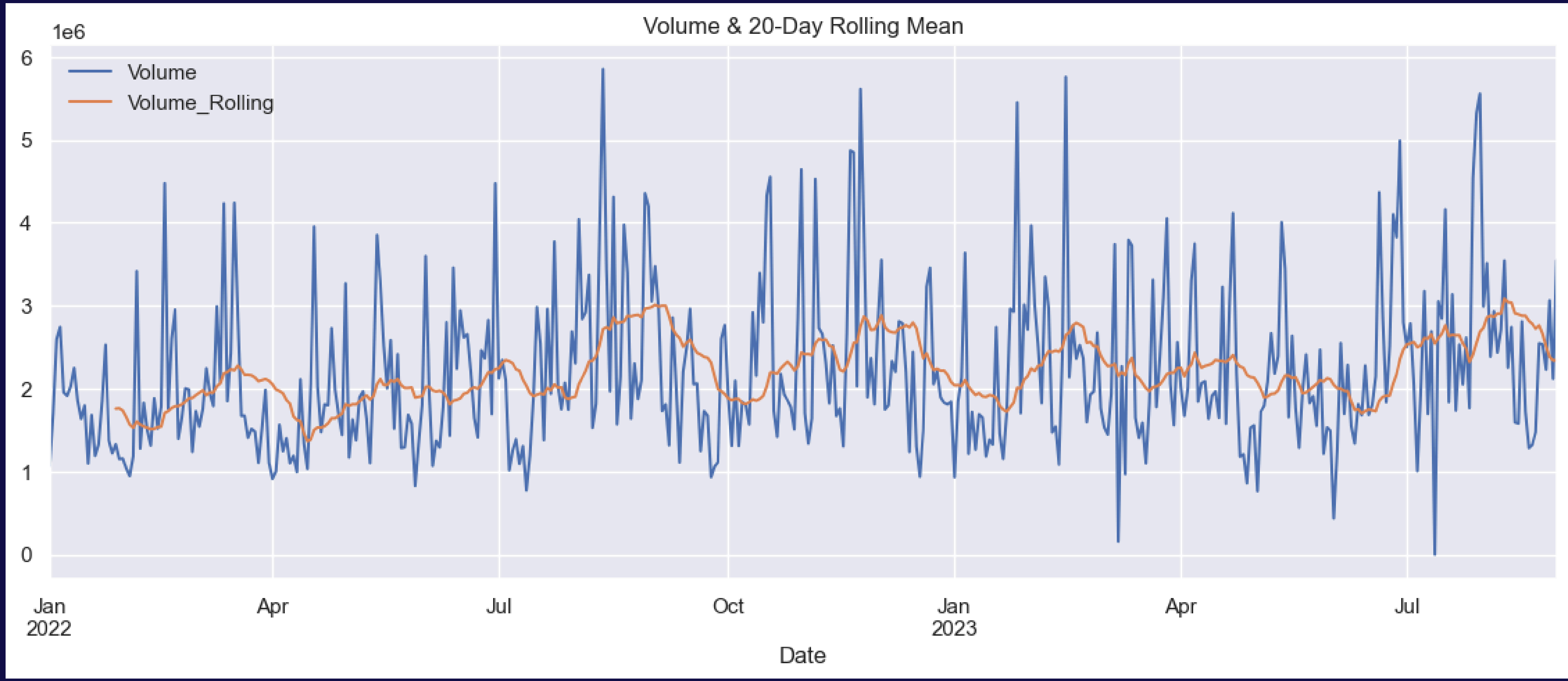


ACF & PACF (Autocorrelation)

ACF & PACF

- ACF: Strong autocorrelation across lags especially for the first 20 lags, justifying lag-based features.

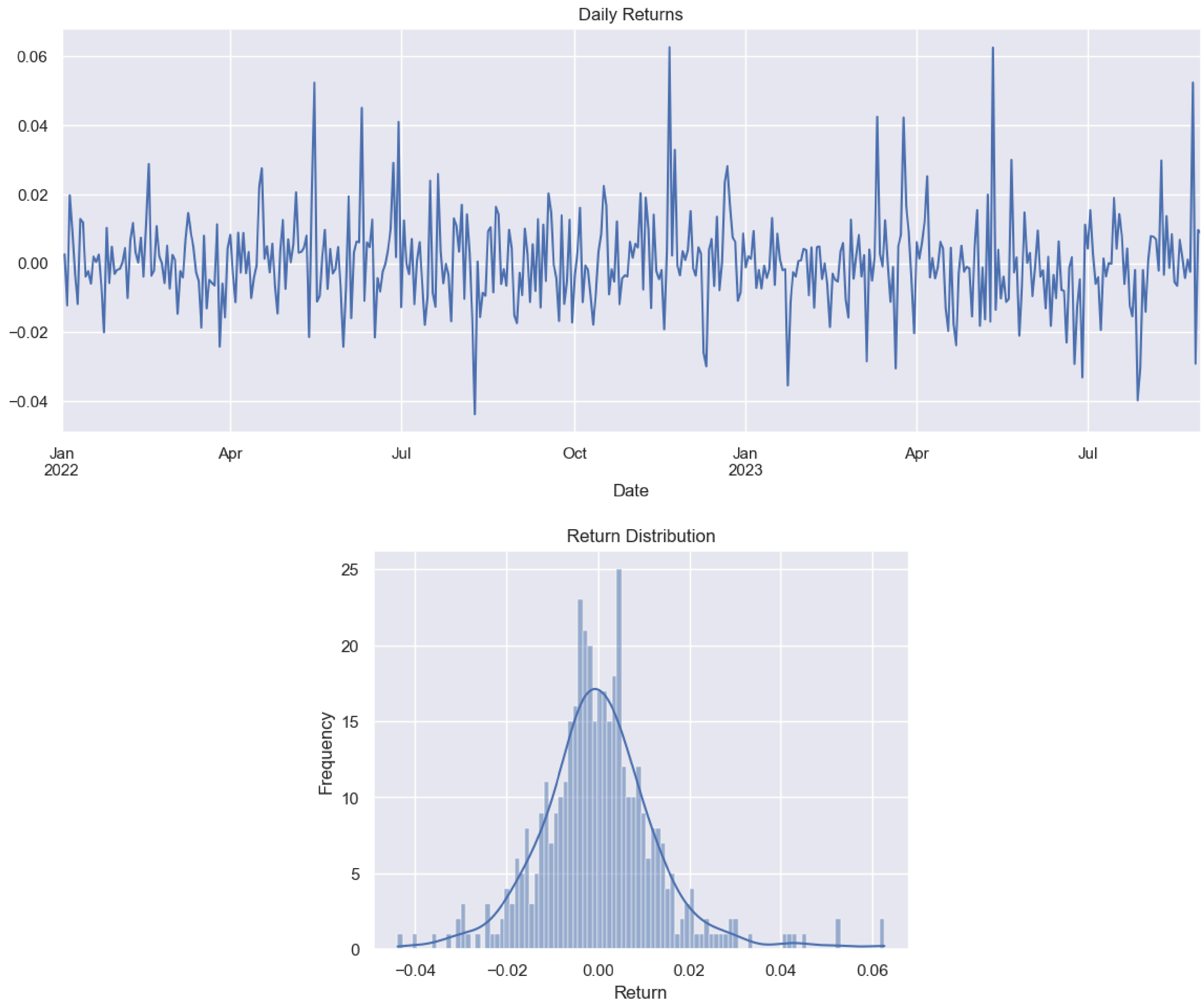
- PACF: Sharp drop after lag-1, validating a potential AR(1) or ARIMA model base.



Volume Analysis

Volume Analysis

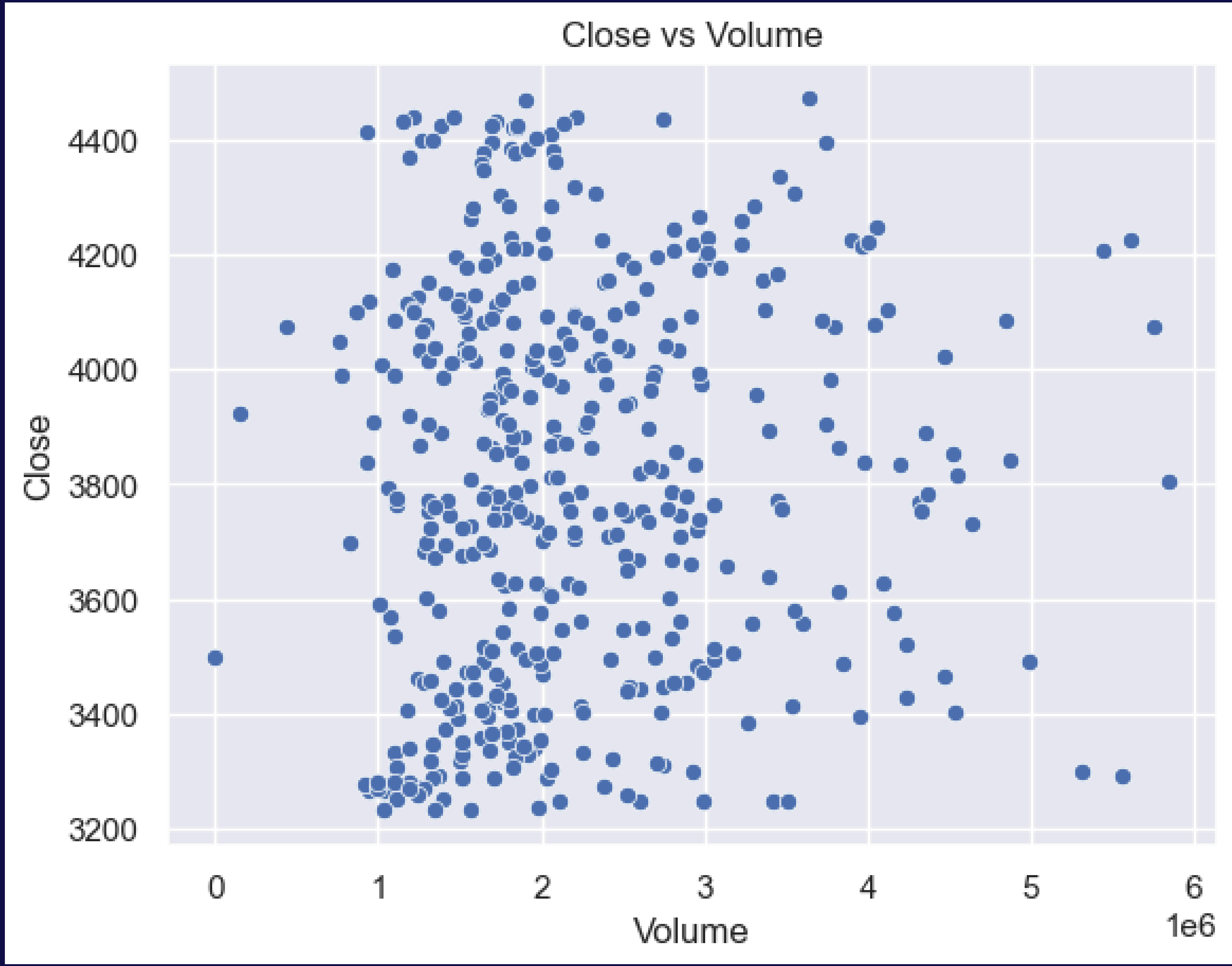
- Volume is highly volatile, but the rolling average smooths out trading trends.
- Volume spikes may correlate with price action or major market events.



Daily Returns & Distribution

Daily Returns & Distribution

- Returns are mostly centered around 0, with a slight right skew.
- Normal distribution assumption is mostly valid, supporting probabilistic models.
- Presence of fat tails—extreme returns should be handled using robust metrics.



Close Price vs Volume (Scatter Plot)

Close Price vs Volume

- Observation: There is no clear linear correlation between closing price and trading volume.
- Inference: Volume does not directly dictate price levels. However, outlier volume spikes might be associated with strong price moves due to news or events (but not consistently).
- Confirms earlier correlation matrix result of weak correlation between Volume and Close.

Cumulative Returns

Cumulative Returns

- Observation: TCS stock has experienced long-term uptrends, peaking around Jan 2023, followed by a consistent downward trend.
- Inference: A buy-and-hold strategy between Jan 2022 and Jan 2023 would have yielded high returns. Post that, investors faced losses.
- Suggests trend reversal models or profit-booking points are crucial for long-term strategy.

RSI Zones Count

```python
overbought = (df['RSI'] > 70).sum()
oversold = (df['RSI'] < 30).sum()
print(f"Overbought days: {overbought}, Oversold days: {oversold}")
```

```
Overbought days: 68, Oversold days: 70
```

RSI Zones Count

RSI Zones

Overbought Days: 68

Oversold Days: 70

Inference:

- RSI crossed thresholds of 70 and 30 nearly equally, showing a volatile stock.
- RSI can be an effective signal for swing trading and identifying potential reversals.

MACD Crossover

MACD Crossovers Count: 38

Inference:

- A moderate number of crossover signals.
- MACD remains a strong trend-following tool, especially for confirming entry/exit during persistent moves.

MACD Crossover Signal Count

```python
df['MACD_Cross'] = df['MACD'] > df['Signal_Line']
crossovers = df['MACD_Cross'].astype(int).diff().fillna(0).abs().sum()
print(f"MACD crossover signals: {int(crossovers)}")
```

```
MACD crossover signals: 38
```

MACD Crossover Signal Count

Model Comparison

RANDOM FOREST & XGBOOST

Both models performed exceptionally well for short-term (1D) forecasts, closely tracking price movements.

- 1D Random Forest: RMSE: 2.65, R^2: 0.95
- 1D XGBoost: RMSE: 2.74, R^2: 0.94

However, performance degraded over mid- (5D, 1M) and long-term (6M) horizons:

- Mid-term predictions began to flatten, unable to capture trend shifts
- Long-term forecasts showed high error and negative R^2, reflecting poor generalization and loss of time awareness

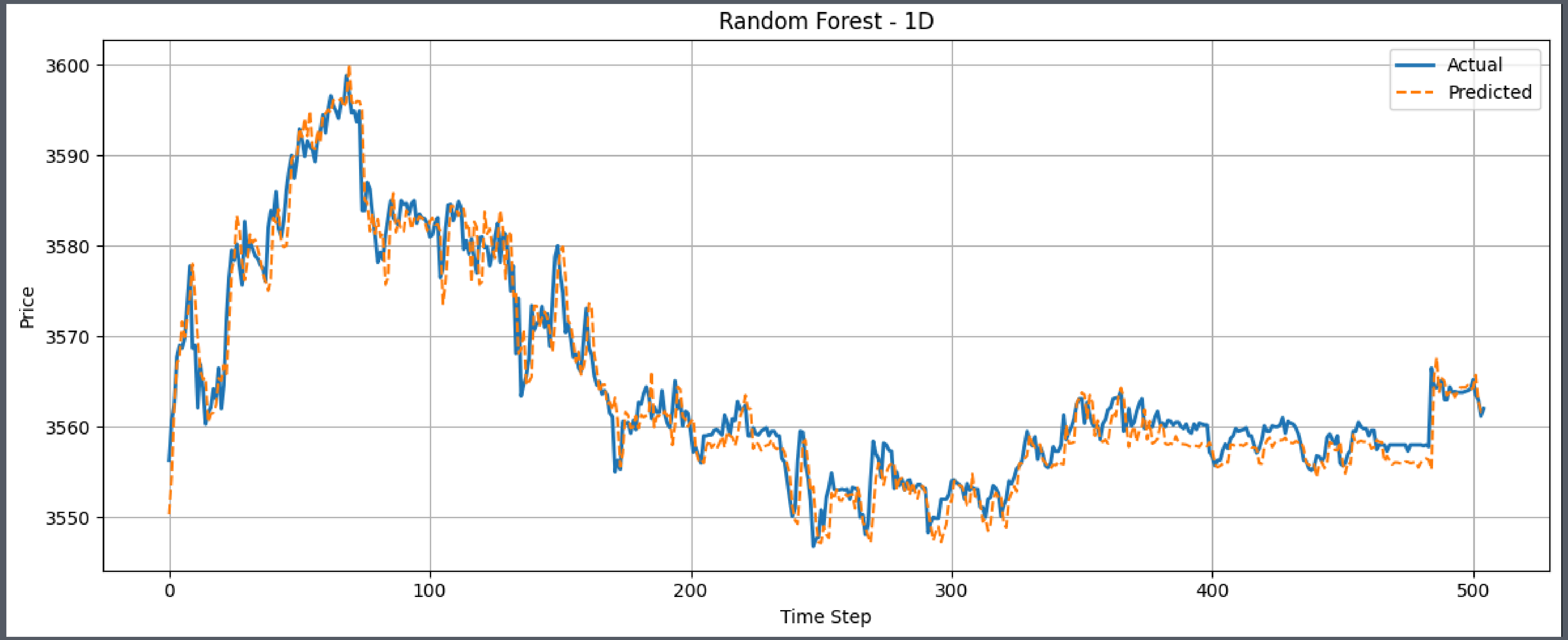
Tree models excel in daily predictions but are limited by their non-sequential architecture when stretched over time.

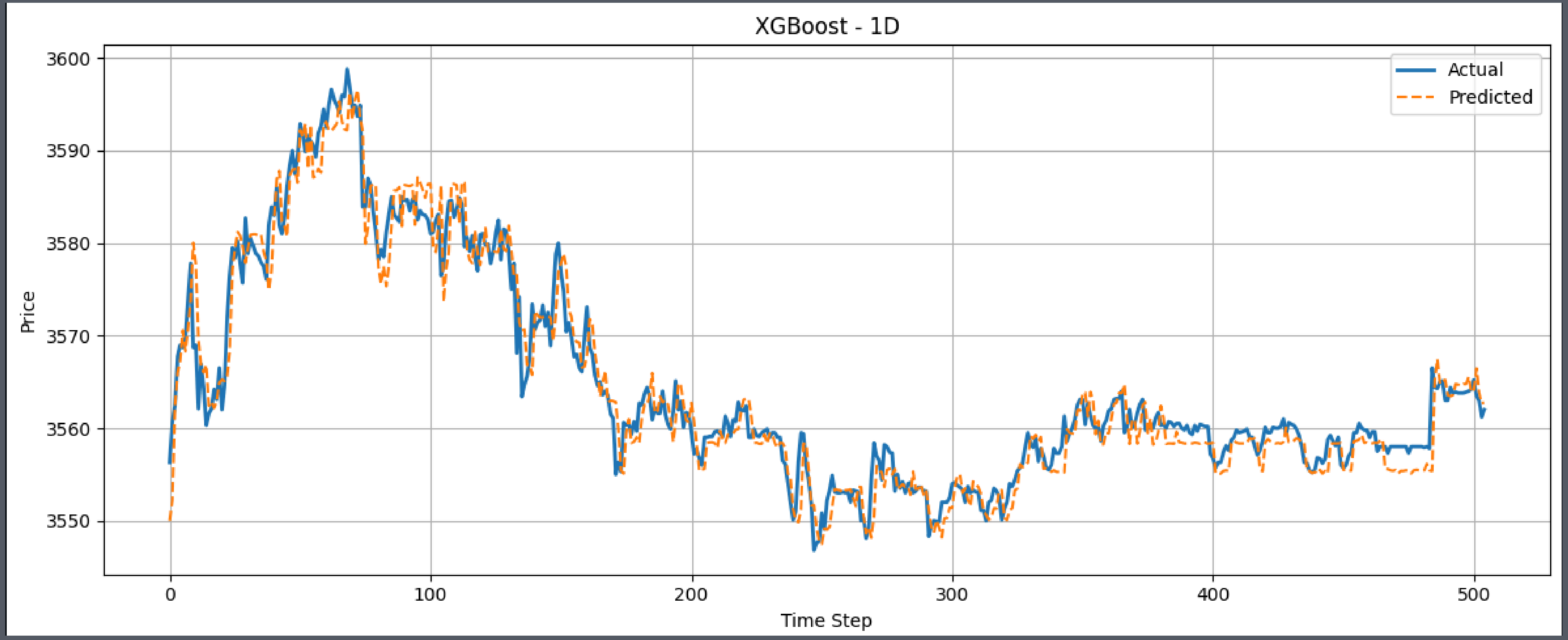
LSTM (DEEP LEARNING)

LSTM models leveraged time-series memory to capture underlying momentum and trend patterns, outperforming tree models on longer sequences:

- 1Y LSTM: RMSE: 151.71, R^2: -0.59
- 5Y LSTM: RMSE: 445.78, R^2: -1.04

While effective in modeling sequential data, LSTMs struggled with volatility and often produced over-smoothed outputs, especially on sharp reversals or during high market noise.

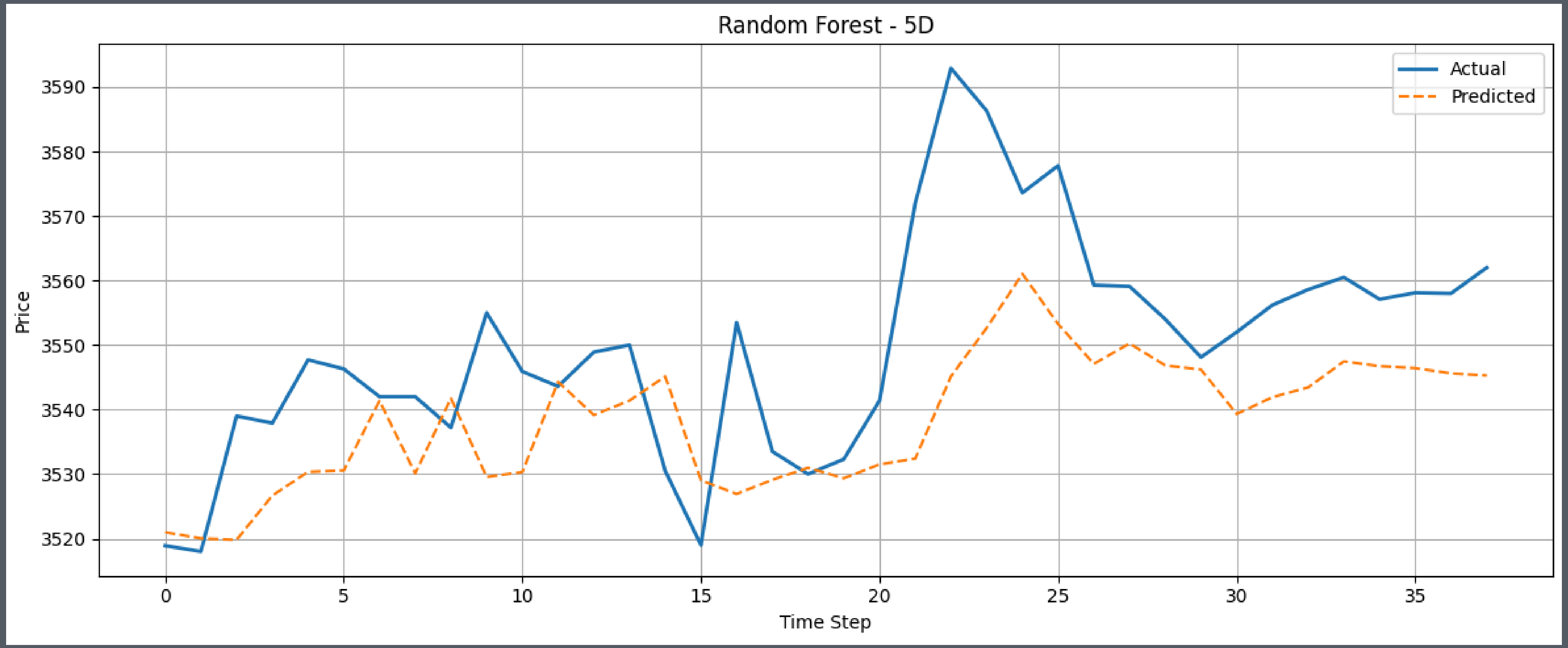


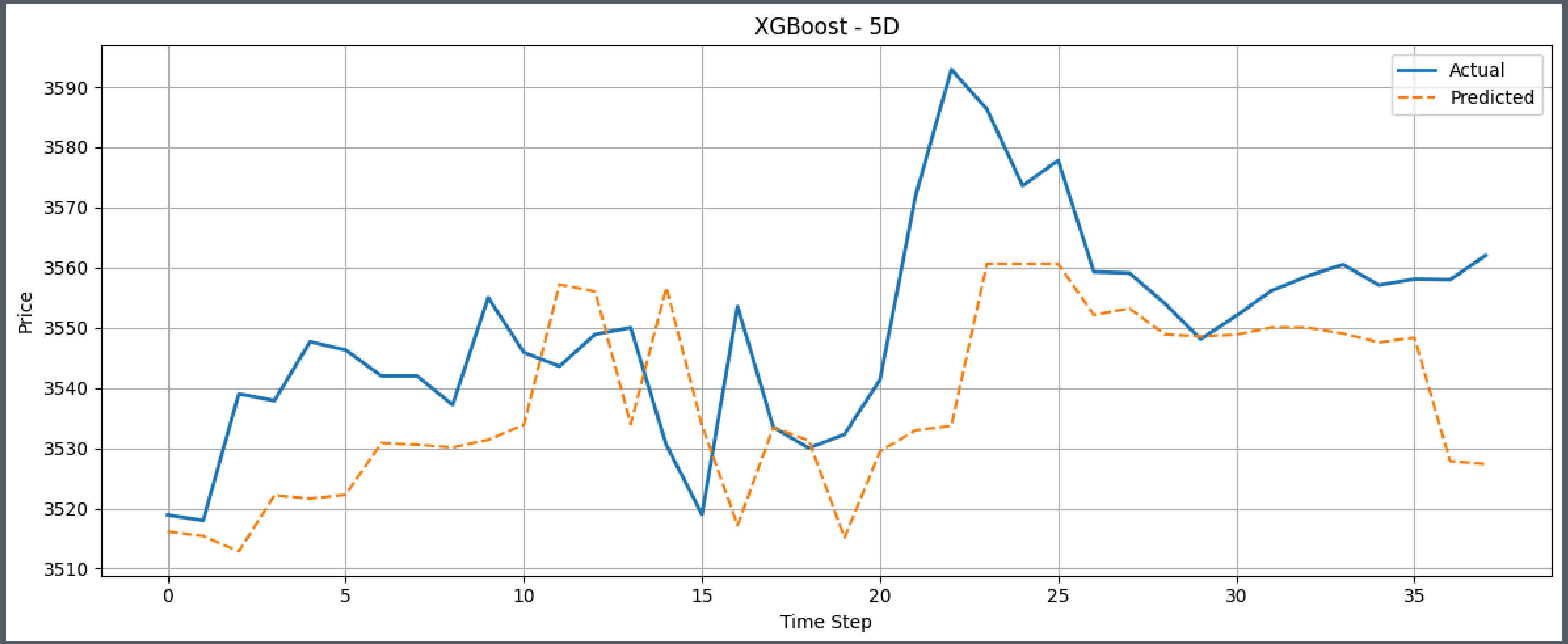


RF and Xgboost 1D

RF and Xgboost 1D

- Random Forest achieved an R^2 of 0.95 and RMSE of 2.65, closely tracking actual price movements.
- XGBoost also performed well with R^2 of 0.94 and RMSE of 2.74, showing slightly more lag on sharp turns.
- These tree-based models effectively captured recent patterns and momentum, making them well-suited for daily or intraday forecasting.





RF and Xgboost 5D

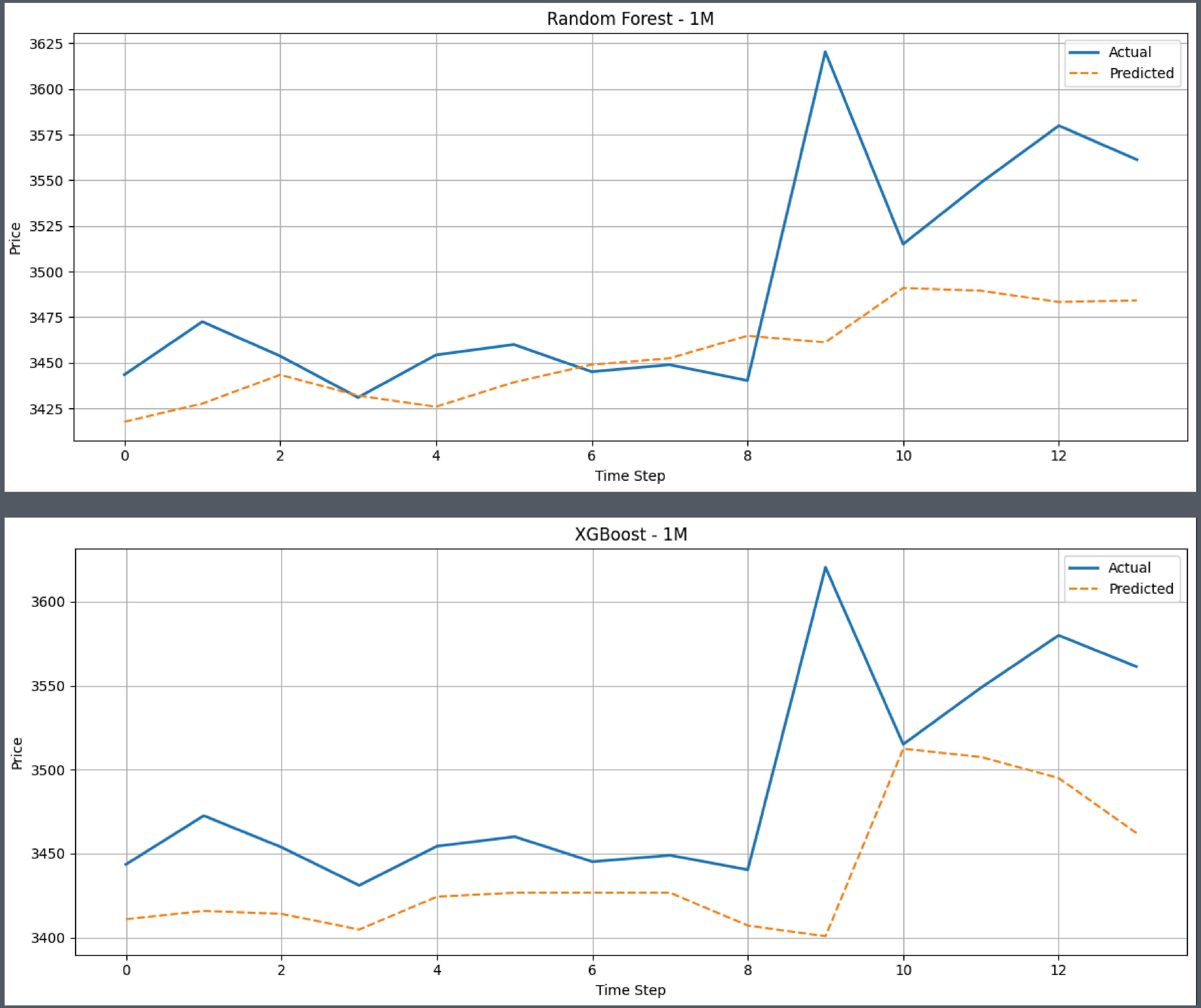
RF and Xgboost 5D

5 days, both Random Forest and XGBoost models show a noticeable drop in accuracy:

- Random Forest: R^2 = -0.01, RMSE = 16.93
- XGBoost: R^2 = -0.41, RMSE = 19.98

Observation: Predictions become flatter and diverge from actual prices, indicating difficulty in capturing short-term volatility.

Inference: Tree-based models begin to lose pattern sensitivity in mid-term horizons due to their lack of temporal awareness.



RF and Xgboost 1M

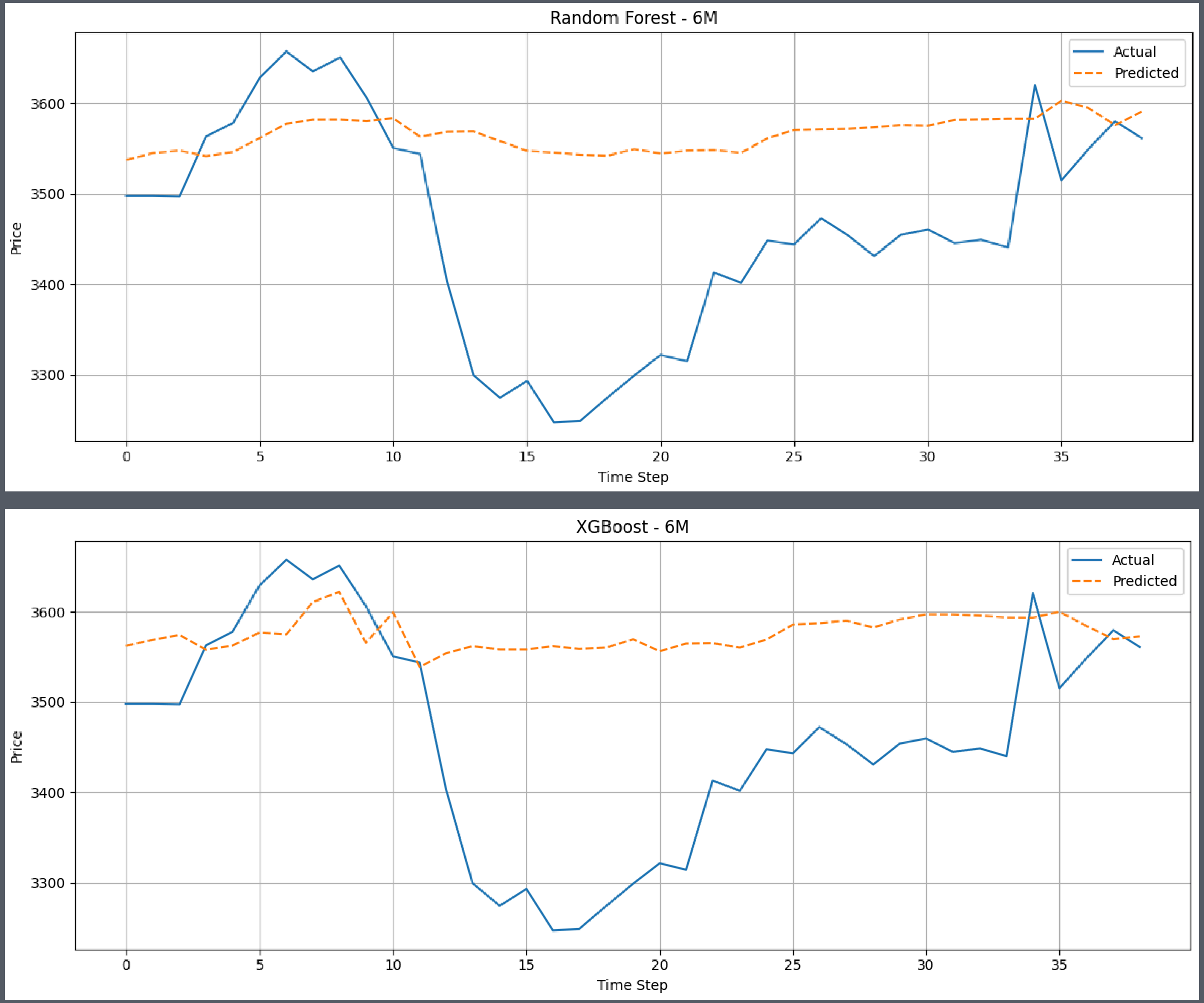
RF and Xgboost 1M

For mid-range forecasting (1M horizon), both Random Forest and XGBoost show limited accuracy. My models fail to capture the sharp upward spikes seen in actual price trends, resulting in low or negative R^2 values.

- Random Forest yields an R^2 of 0.01, struggling with price volatility.

- XGBoost performs worse with an R^2 of -0.56, underpredicting during high jumps.

These results highlight the limitation of tree-based models in adapting to medium-term fluctuations and reinforce the need for time-aware architectures in volatile datasets.



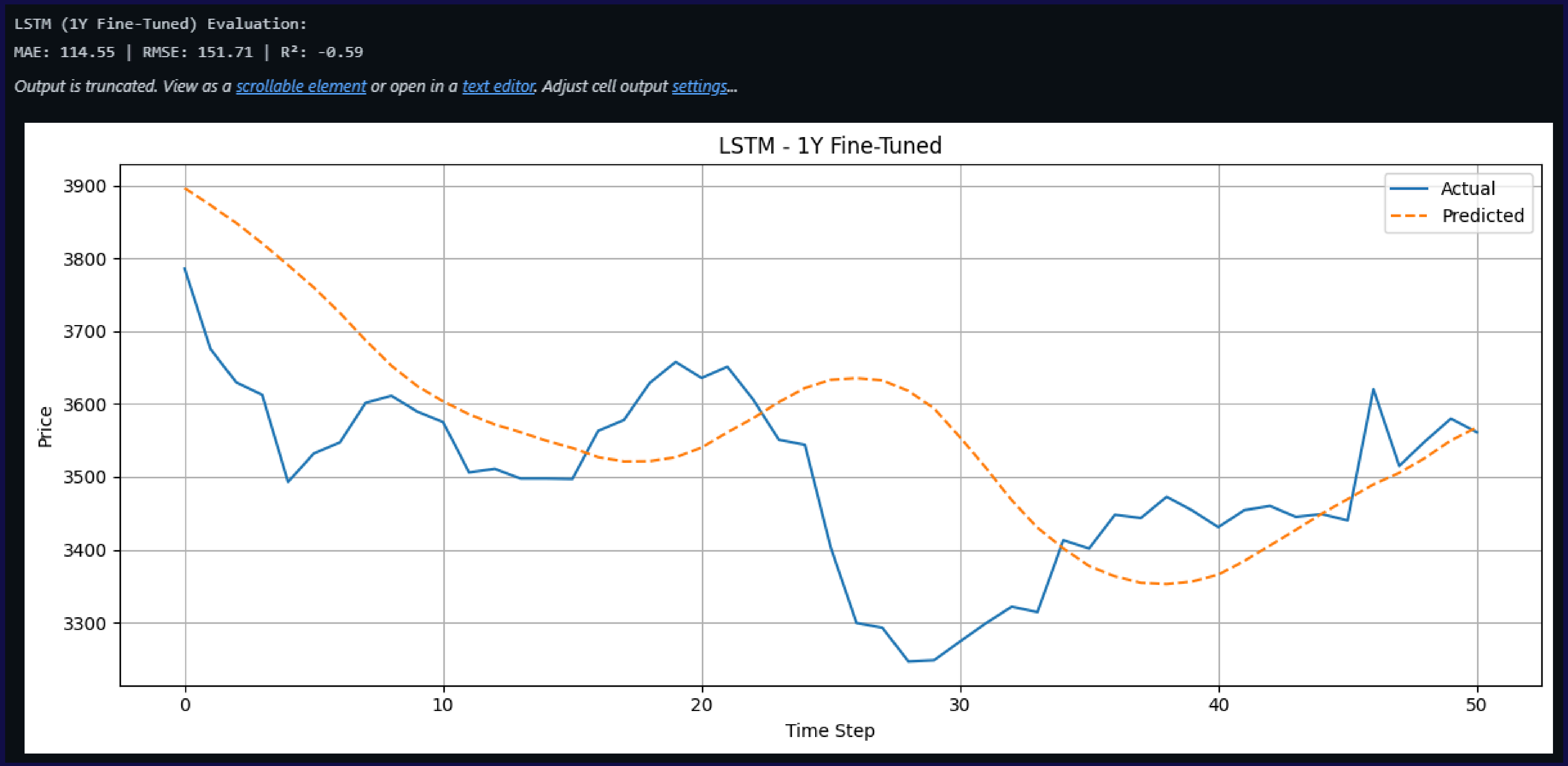
RF and Xgboost 6M

RF and Xgboost 6M

Long-Term (6 Months)

- Random Forest (R^2: -0.63) and XGBoost (R^2: -0.81) both fail to capture long-term fluctuations.
- Predictions flatten significantly, diverging from actual trend reversals and spikes.

Conclusion: Neither model is suitable for investment-level long-term forecasting without integrating sequential learning.



LSTM – 1 Year Forecast (Fine-Tuned)

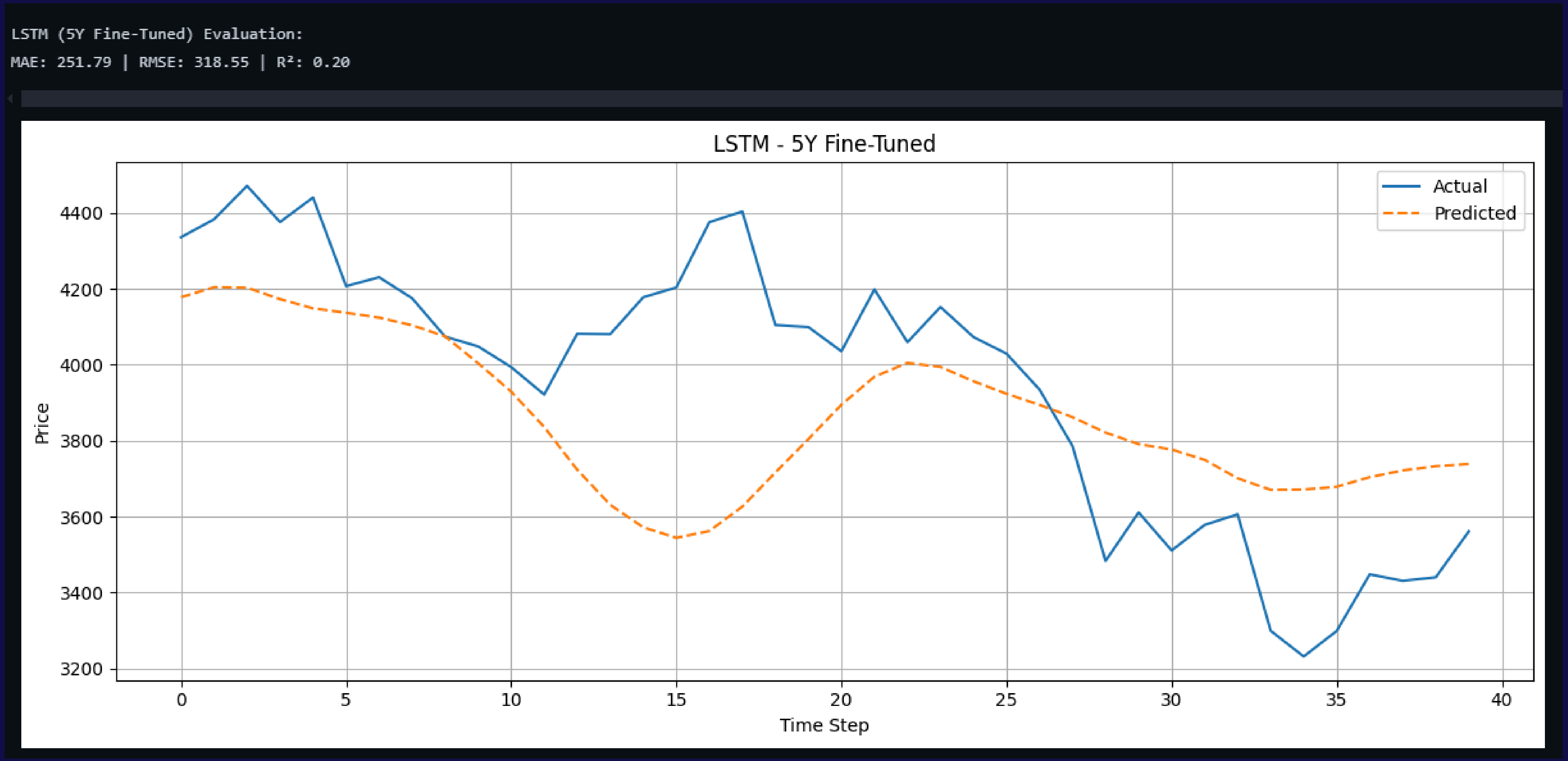
LSTM – 1Y

The LSTM model demonstrates its ability to learn sequential patterns over a long horizon but struggles to follow sharp market reversals. With an R^2 of -0.59 and RMSE of 151.71, the predictions appear over-smoothed, failing to capture the volatility of the actual price movements.

Strength: Captures overall trend direction

Limitation: Misses short-term spikes and dips

Use Case Fit: Moderate for trend analysis; not reliable for precise long-term price prediction without enhanced tuning or hybrid modeling.



LSTM: Actual vs Predicted Close Price

LSTM

The fine-tuned LSTM model better captures long-term trends with moderate accuracy (R^2 = 0.20, RMSE = 318.55). It aligns with overall market direction but still smooths out sharp volatility. Ideal for high-level forecasting, though further tuning or external features could enhance precision.

- The model tracks trend direction but underreacts to rapid market changes.
- LSTM benefits from temporal patterns but may require enriched features to model volatility.
- Short-term dips and spikes are not reflected with high fidelity.

Key Takeaways

STRONG SHORT-TERM MODEL PERFORMANCE

Random Forest and XGBoost performed exceptionally well for 1-day predictions, closely tracking price with high accuracy. Ideal for intraday or daily trading strategies.

MID-TERM PREDICTION CHALLENGES

Performance of tree-based models dropped significantly over 5D and 1M periods. Models failed to capture trend reversals or volatility, leading to flattened predictions and negative R^2.

LSTM CAPTURES TEMPORAL PATTERNS

LSTM outperformed classical models in modeling time-dependent momentum and sequential price behavior. However, it struggled during sharp reversals and volatile spikes.

LONG-TERM FORECASTS BREAK DOWN

Across 6M to 5Y ranges, all models showed high error and instability. LSTM outputs became overly smoothed, while tree models lost predictive relevance entirely.

TECHNICAL INDICATOR IMPACT

Indicators like RSI, MACD, MA20/MA50, and Bollinger Bands enhanced feature richness. Their inclusion boosted model understanding of market momentum and trend shifts.

NEED FOR HYBRID & ADAPTIVE MODELS

Long-term accuracy requires hybrid approaches (e.g., ARIMA + LSTM or CNN-LSTM) and retraining pipelines that adapt to regime changes and reduce drift.(Next Time)

TCS Stock Dashboard App (app.py)

This interactive web app was built using Streamlit to visualize and analyze TCS stock price forecasts based on various ML models.

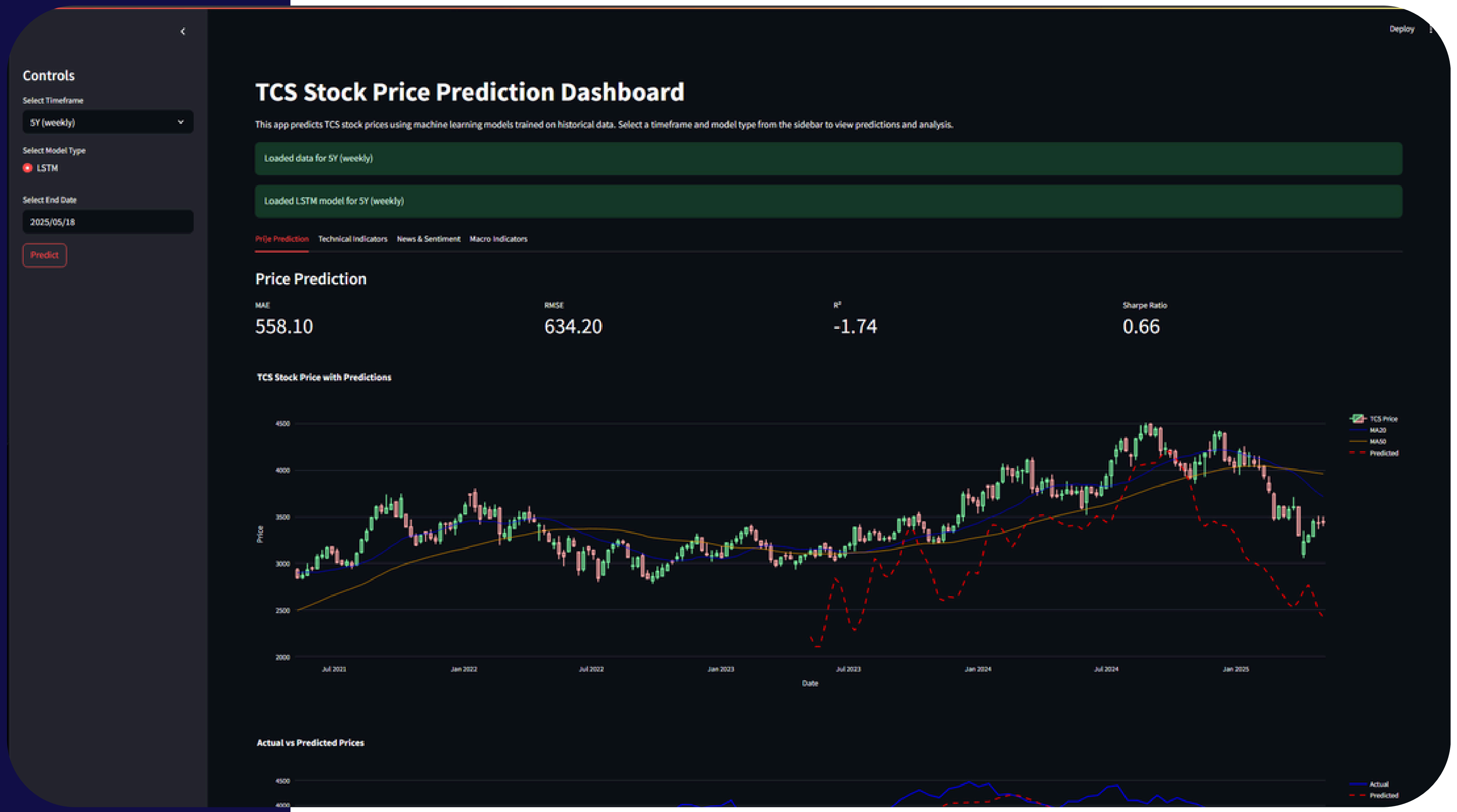
The interface allows users to:

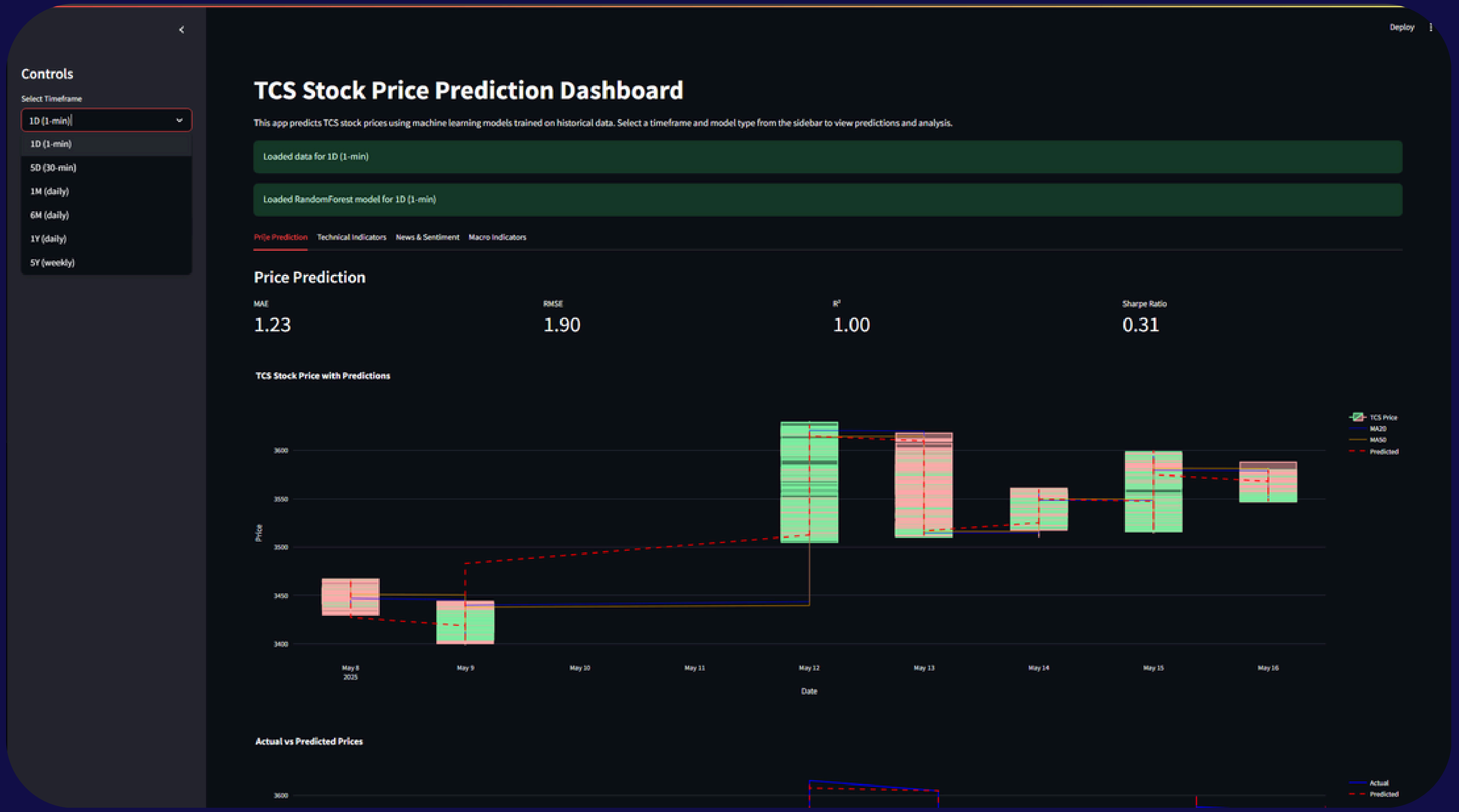
- Select timeframe: Choose between 1D (1-min), 5D (30-min), 1M, 6M, 1Y, or 5Y intervals.
- Select model type: Toggle between trained Random Forest, XGBoost, or LSTM models.
- Pick end date: Adjust the forecast window dynamically.

Features Included:

- Price prediction visualization with actual vs. predicted comparison.
- Technical indicator overlays like MA20, MA50, RSI, and MACD.
- Buy/Sell signal generation using EMA crossover logic.
- Error metrics (MAE, RMSE, R^2, Sharpe Ratio) displayed for easy model evaluation.
- Tabs for analysis: Navigate through price prediction, technicals, sentiment, and macro factors for deeper insights.

This dashboard effectively demonstrates how traditional ML and deep learning models behave across different time horizons, supporting data-driven trading and investment strategies.





Thank You