

The background of the slide is a complex, abstract pattern of light blue and white geometric shapes, including triangles and polygons, connected by thin lines. This pattern is overlaid on a dark blue background. The title text is centered in a large, white, sans-serif font.

Stock Price Forecast

Younes Mahdavi
Oct 2024

Amazon 1997 – 2024: ~ 6900 days



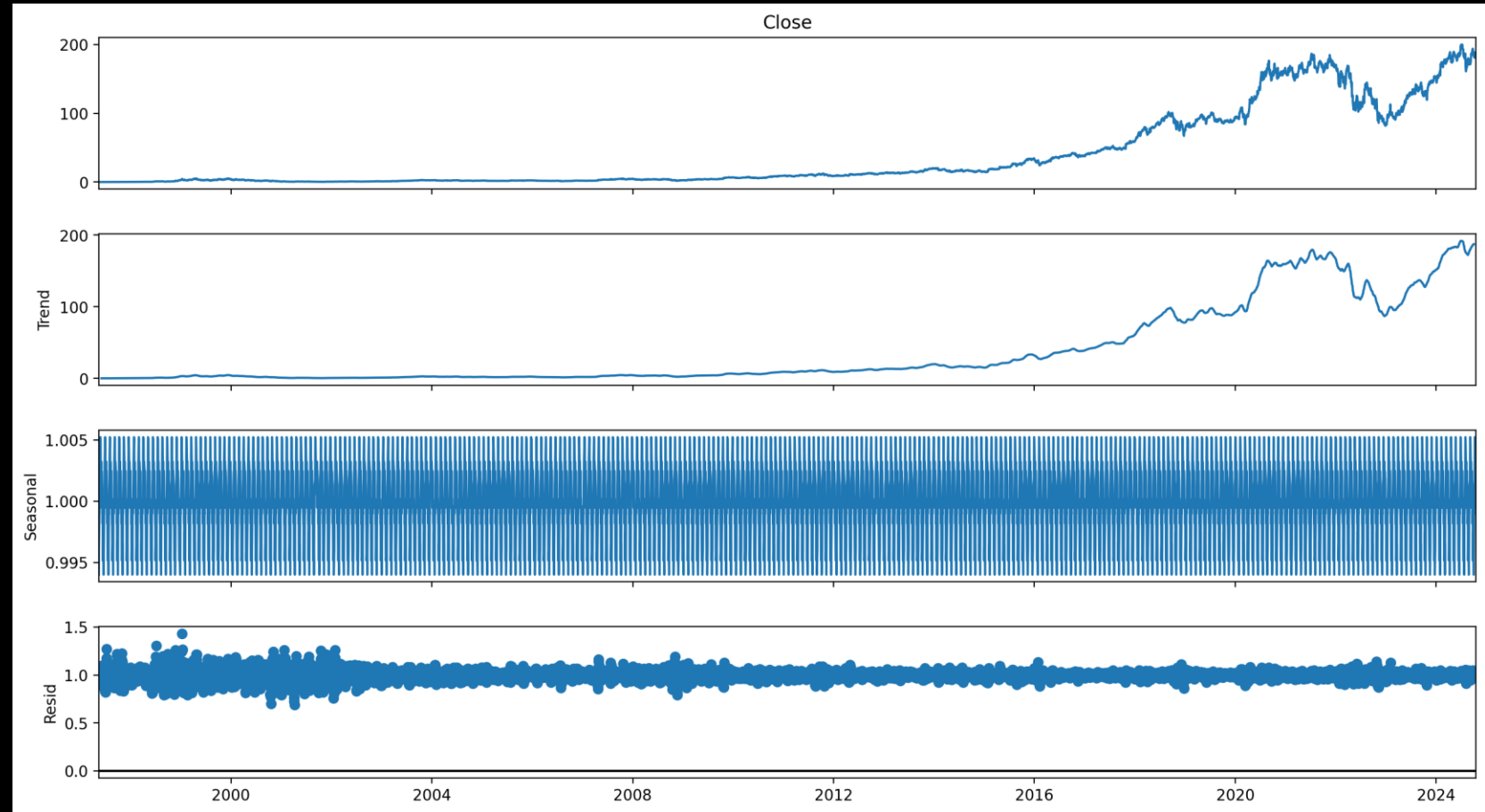
| | Open | High | Low | Close | Volume | Dividends | Stock Splits |
|---------------------------|----------|----------|----------|----------|------------|-----------|--------------|
| Date | | | | | | | |
| 1997-05-15 00:00:00-04:00 | 0.121875 | 0.125000 | 0.096354 | 0.097917 | 1443120000 | 0.0 | 0.0 |
| 1997-05-16 00:00:00-04:00 | 0.098438 | 0.098958 | 0.085417 | 0.086458 | 294000000 | 0.0 | 0.0 |
| 1997-05-19 00:00:00-04:00 | 0.088021 | 0.088542 | 0.081250 | 0.085417 | 122136000 | 0.0 | 0.0 |
| 1997-05-20 00:00:00-04:00 | 0.086458 | 0.087500 | 0.081771 | 0.081771 | 109344000 | 0.0 | 0.0 |
| 1997-05-21 00:00:00-04:00 | 0.081771 | 0.082292 | 0.068750 | 0.071354 | 377064000 | 0.0 | 0.0 |
| ... | ... | ... | ... | ... | ... | ... | ... |

Tesla 2010 – 2024: ~ 3600 days

| | Open | High | Low | Close | Volume | Dividends | Stock Splits |
|---------------------------|----------|----------|----------|----------|-----------|-----------|--------------|
| Date | | | | | | | |
| 2010-06-29 00:00:00-04:00 | 1.266667 | 1.666667 | 1.169333 | 1.592667 | 281494500 | 0.0 | 0.0 |
| 2010-06-30 00:00:00-04:00 | 1.719333 | 2.028000 | 1.553333 | 1.588667 | 257806500 | 0.0 | 0.0 |
| 2010-07-01 00:00:00-04:00 | 1.666667 | 1.728000 | 1.351333 | 1.464000 | 123282000 | 0.0 | 0.0 |
| 2010-07-02 00:00:00-04:00 | 1.533333 | 1.540000 | 1.247333 | 1.280000 | 77097000 | 0.0 | 0.0 |
| 2010-07-06 00:00:00-04:00 | 1.333333 | 1.333333 | 1.055333 | 1.074000 | 103003500 | 0.0 | 0.0 |
| ... | ... | ... | ... | ... | ... | ... | ... |



EDA: Decomposition chart





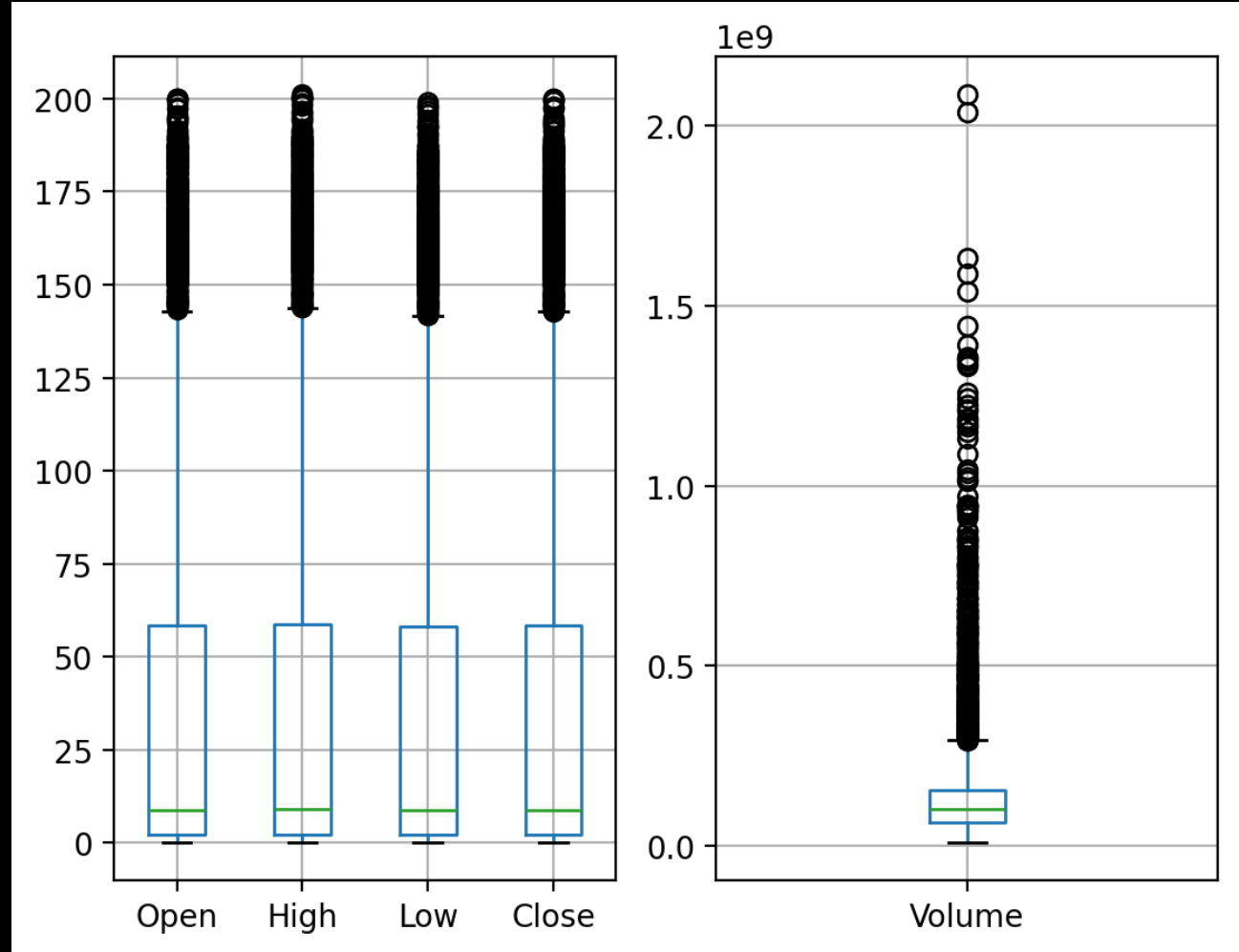
Gap Function

stock_gap()

| | Open | High | Low | Close | Volume | Dividends | Stock Splits |
|---------------------------|--------|---------|---------|---------|-------------|-----------|--------------|
| 2001-09-10 00:00:00-04:00 | 0.420 | 0.4325 | 0.4030 | 0.4315 | 116672000.0 | 0.0 | 0.0 |
| 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 2001-09-17 00:00:00-04:00 | 0.365 | 0.4025 | 0.3525 | 0.3745 | 211470000.0 | 0.0 | 0.0 |
| 2006-12-29 00:00:00-05:00 | 2.003 | 2.0125 | 1.9675 | 1.9730 | 83940000.0 | 0.0 | 0.0 |
| 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 2007-01-03 00:00:00-05:00 | 1.934 | 1.9530 | 1.9025 | 1.9350 | 248102000.0 | 0.0 | 0.0 |
| 2012-10-26 00:00:00-04:00 | 11.430 | 11.9355 | 11.3345 | 11.9120 | 227350000.0 | 0.0 | 0.0 |
| 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 2012-10-31 00:00:00-04:00 | 11.816 | 11.9350 | 11.5250 | 11.6445 | 95952000.0 | 0.0 | 0.0 |

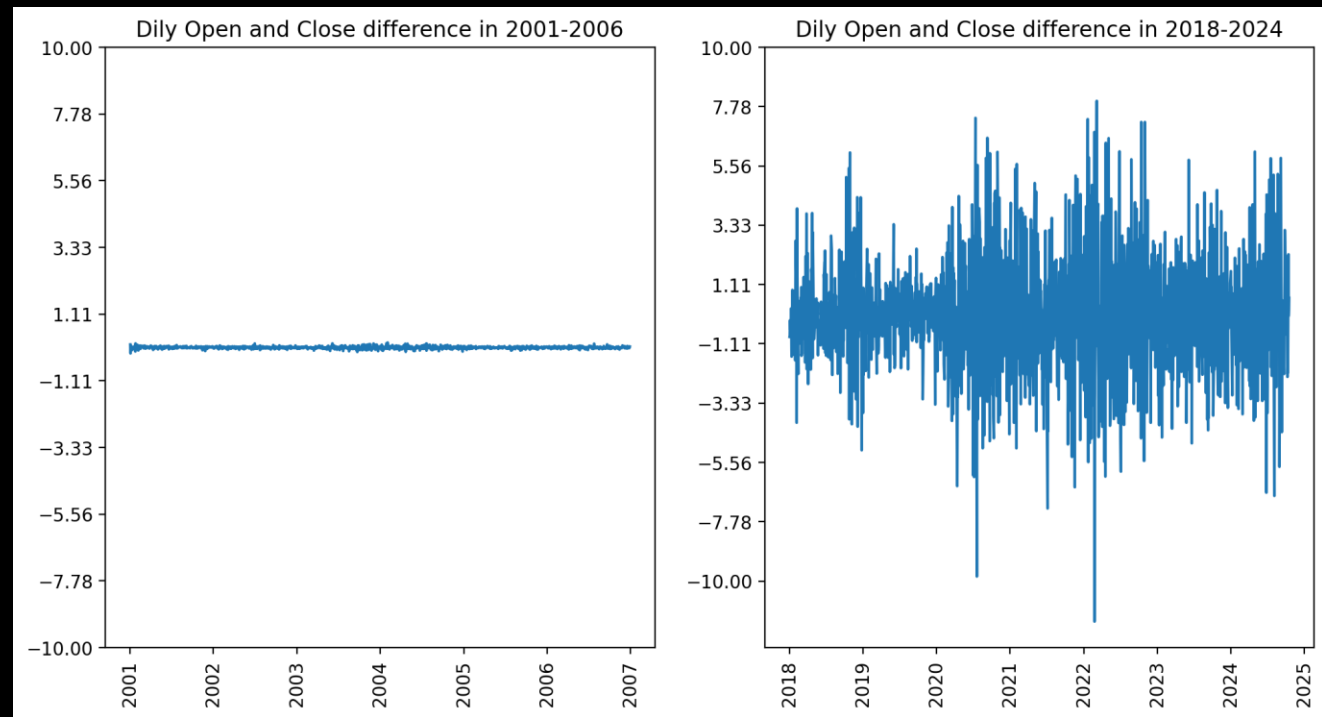


Outliers





Feature selection



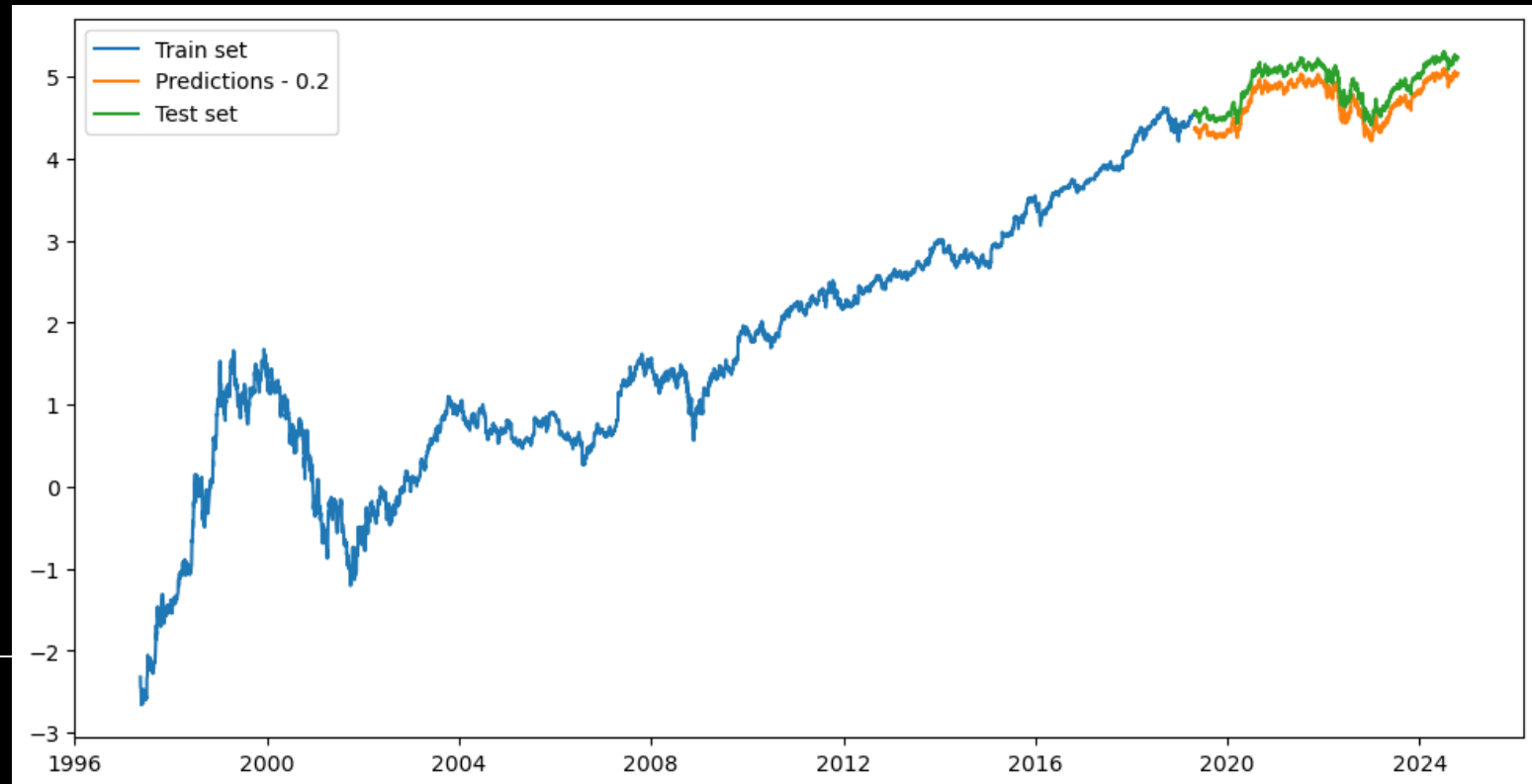
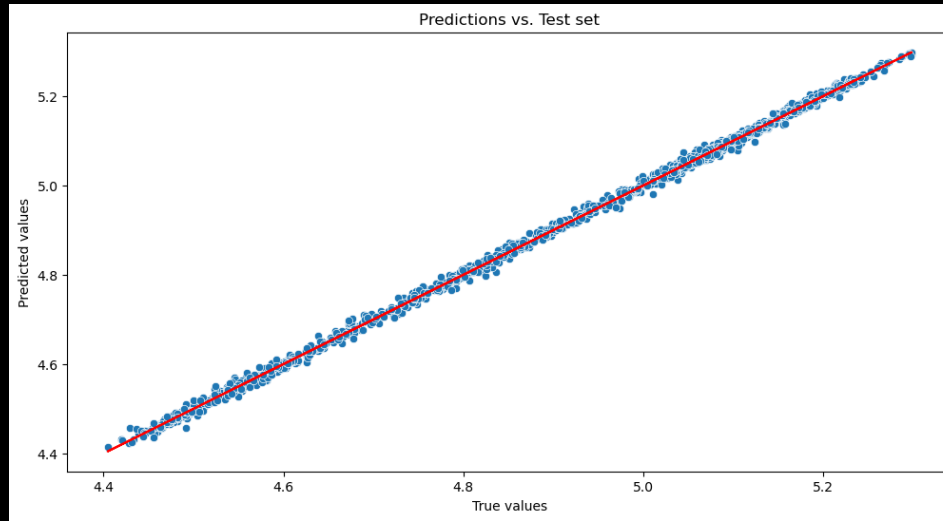


Supervised Learning Models

- 1- Linear Regression, OLS
- 2- Random Forest Regression
- 3- Gradient Boosting Regression



Linear Regression, OLS





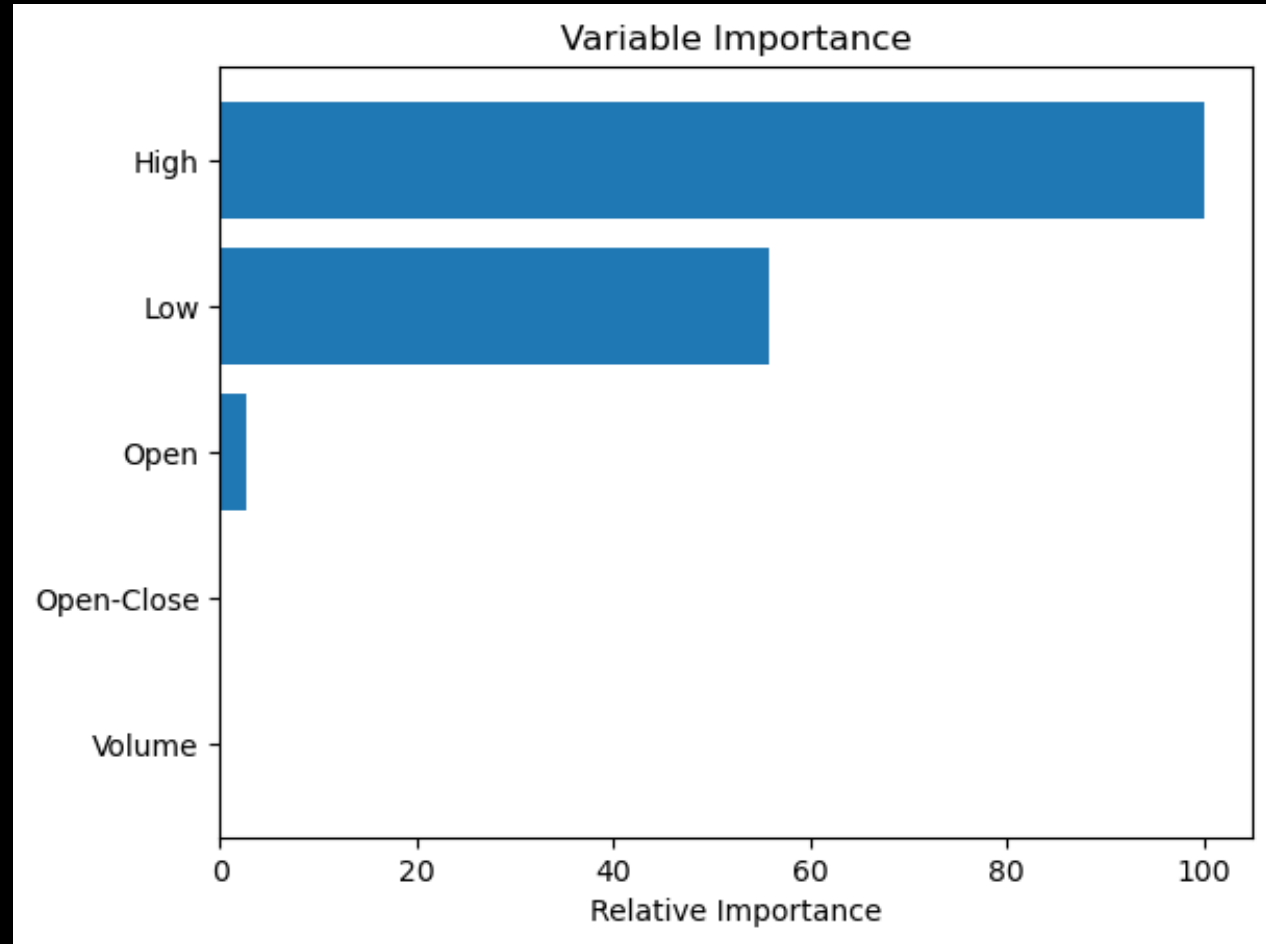
Random Forest Regression



Scores of **10-fold** cross-validation
Accuracy = **0.90**
Standard deviation = **0.12**

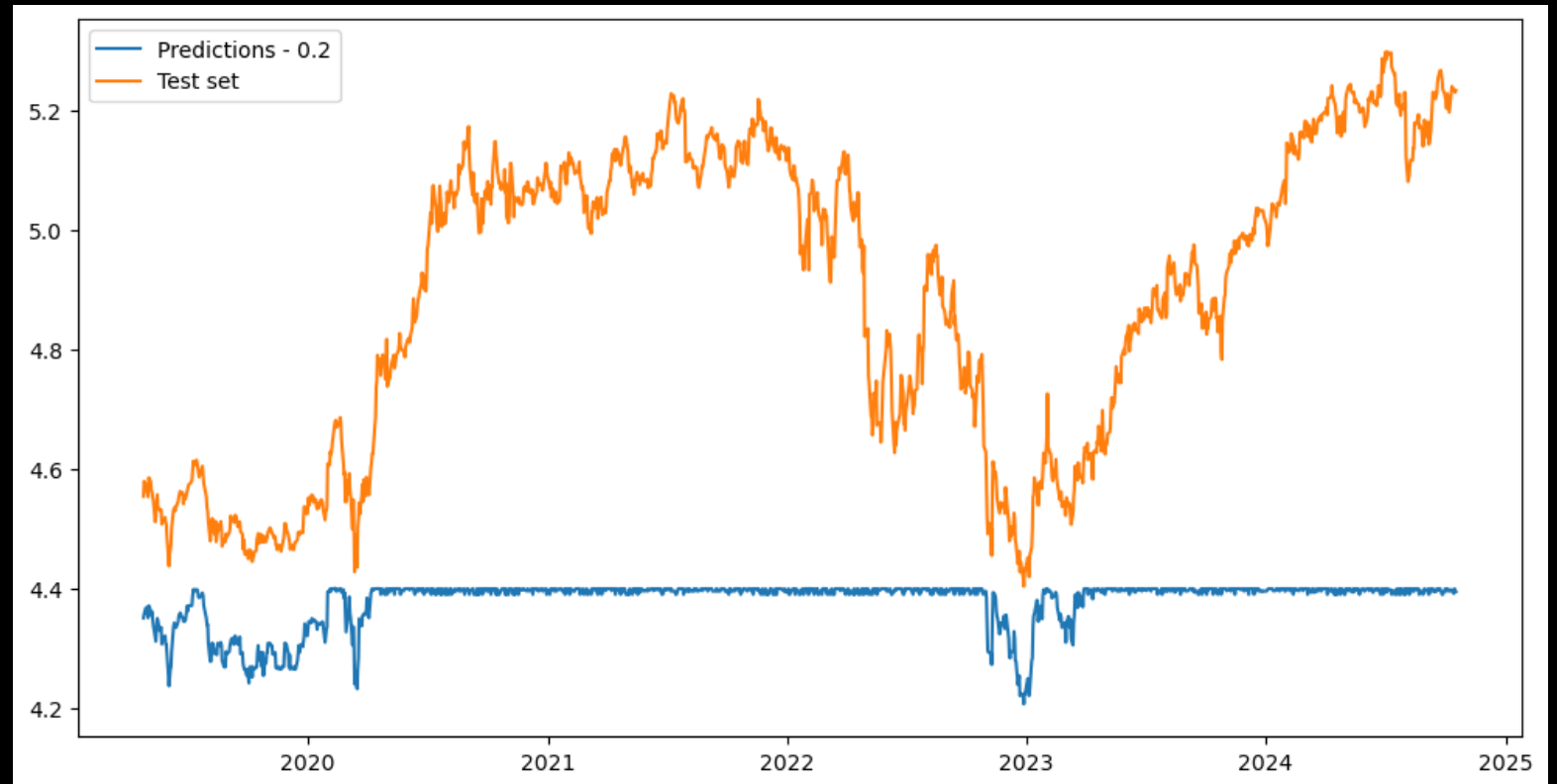


Random Forest Regression





Gradient Boosting Regression



Accuracy of test score = **-1.33**

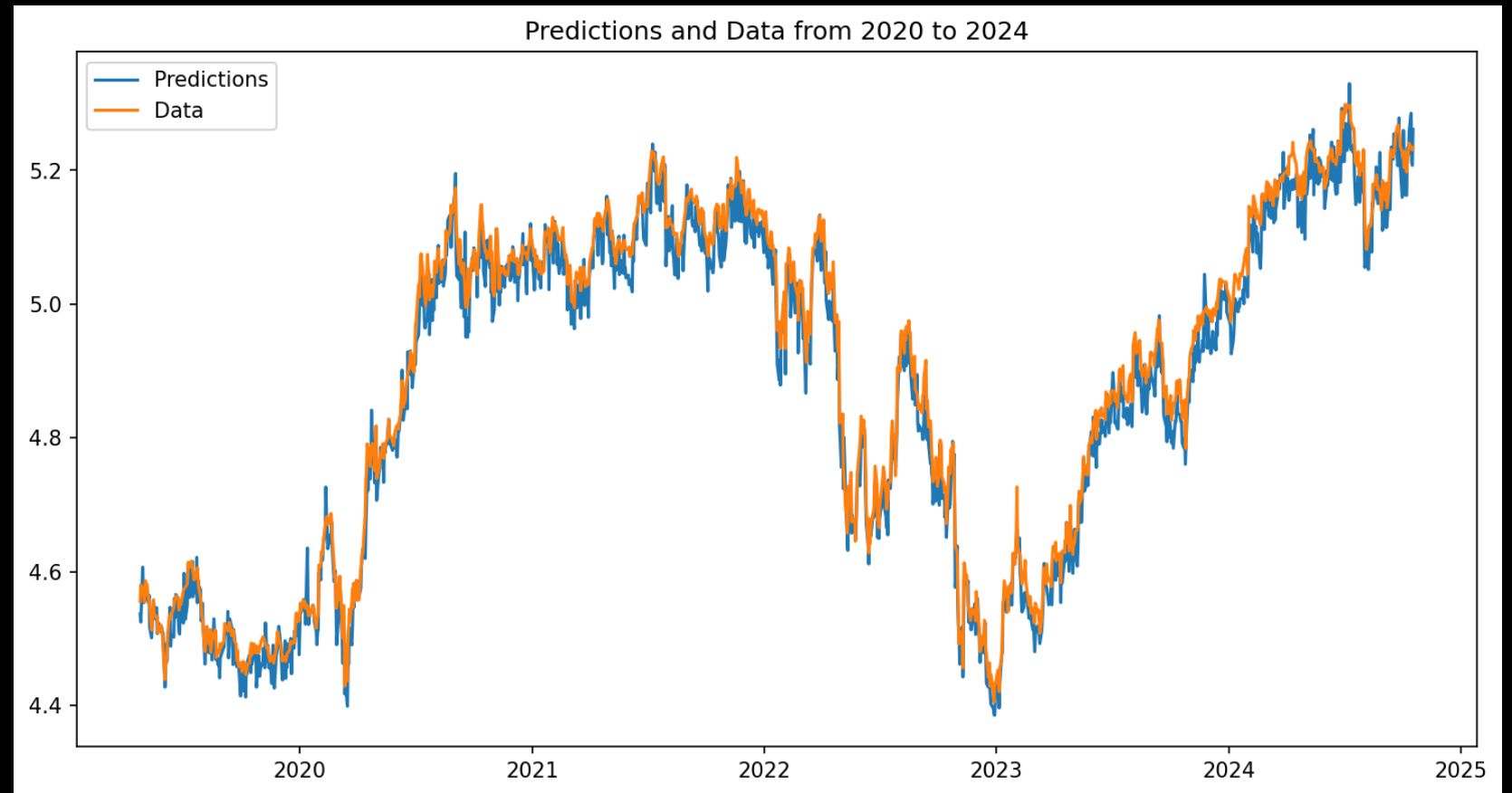
Accuracy of train score = **0.99**



Deep Learning Models



MLP Regression

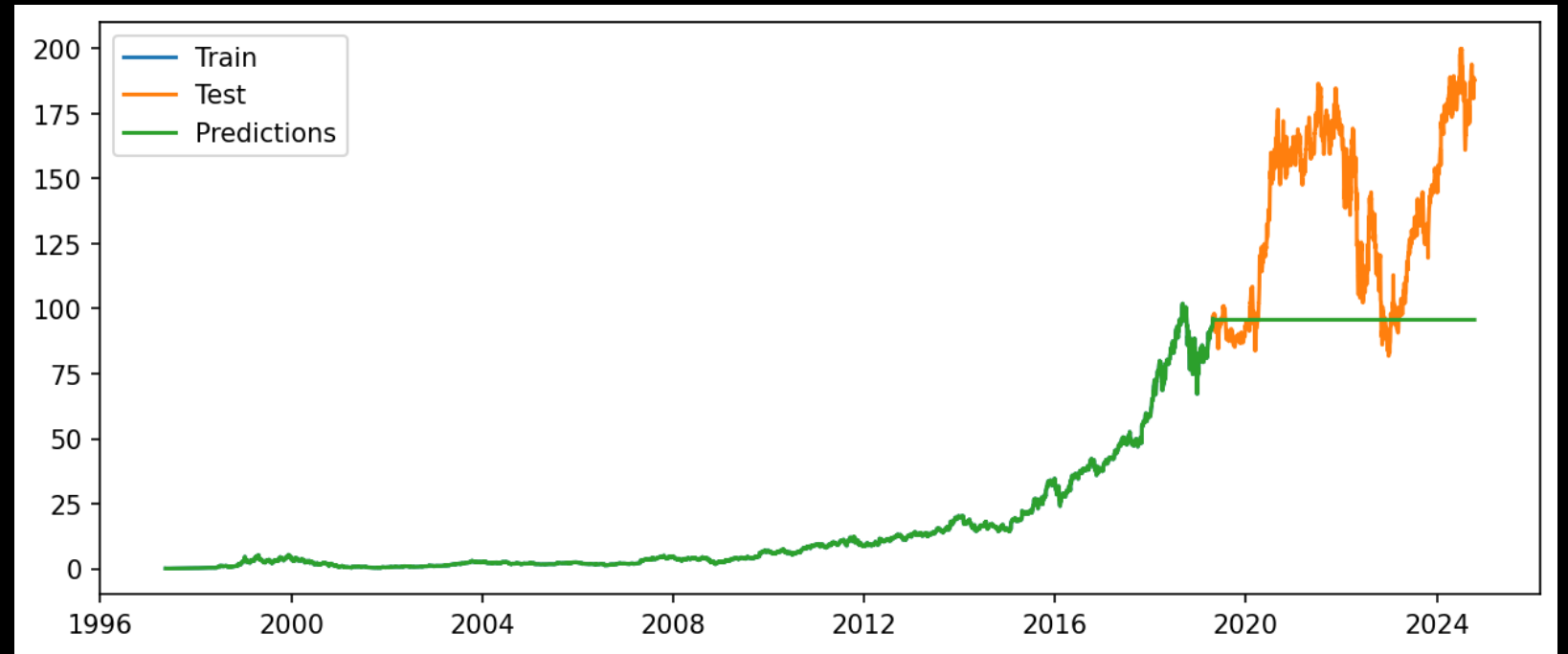


R-squared (training set) = 0.99

R-squared (test set) = 0.98



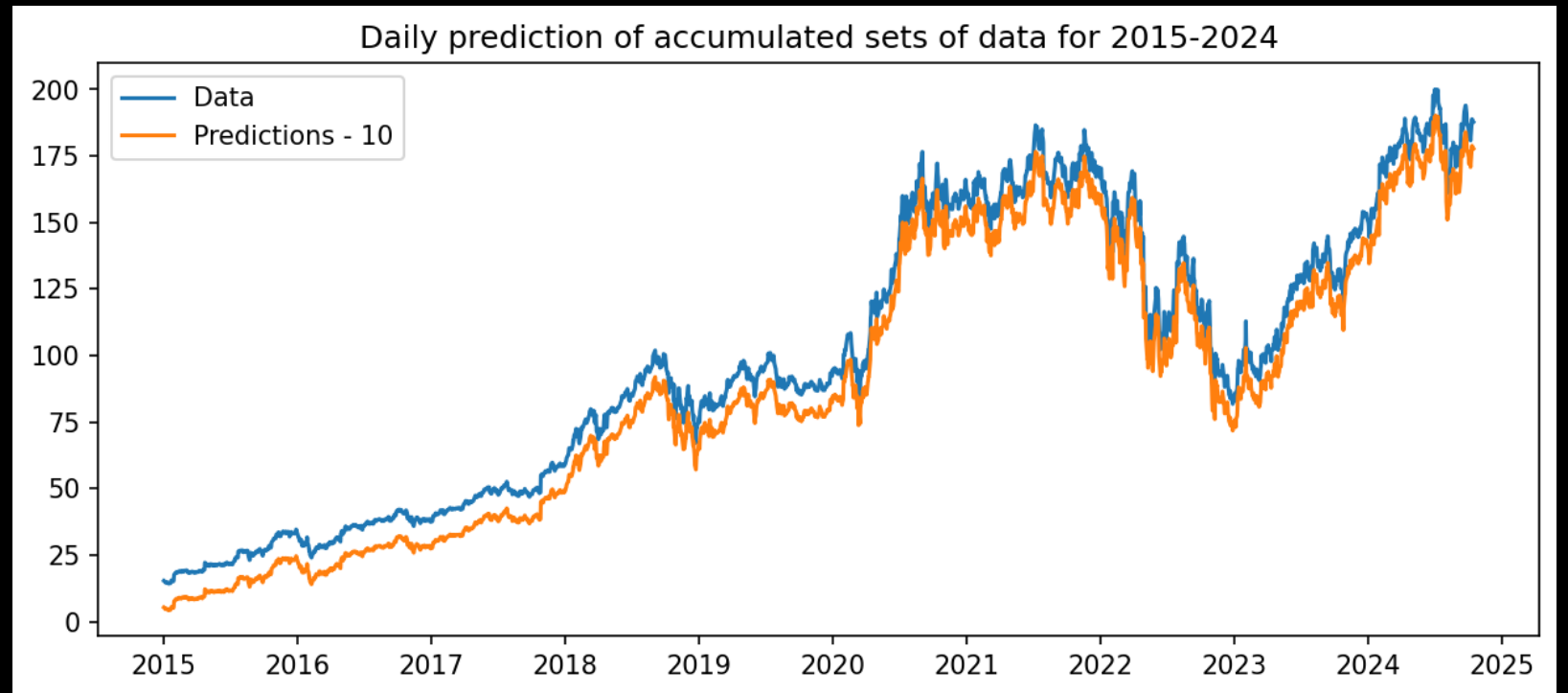
ARIMA Model



ARIMA(1, 1, 1)



ARIMA Model: Accumulative



ARIMA(0, 1, 0)

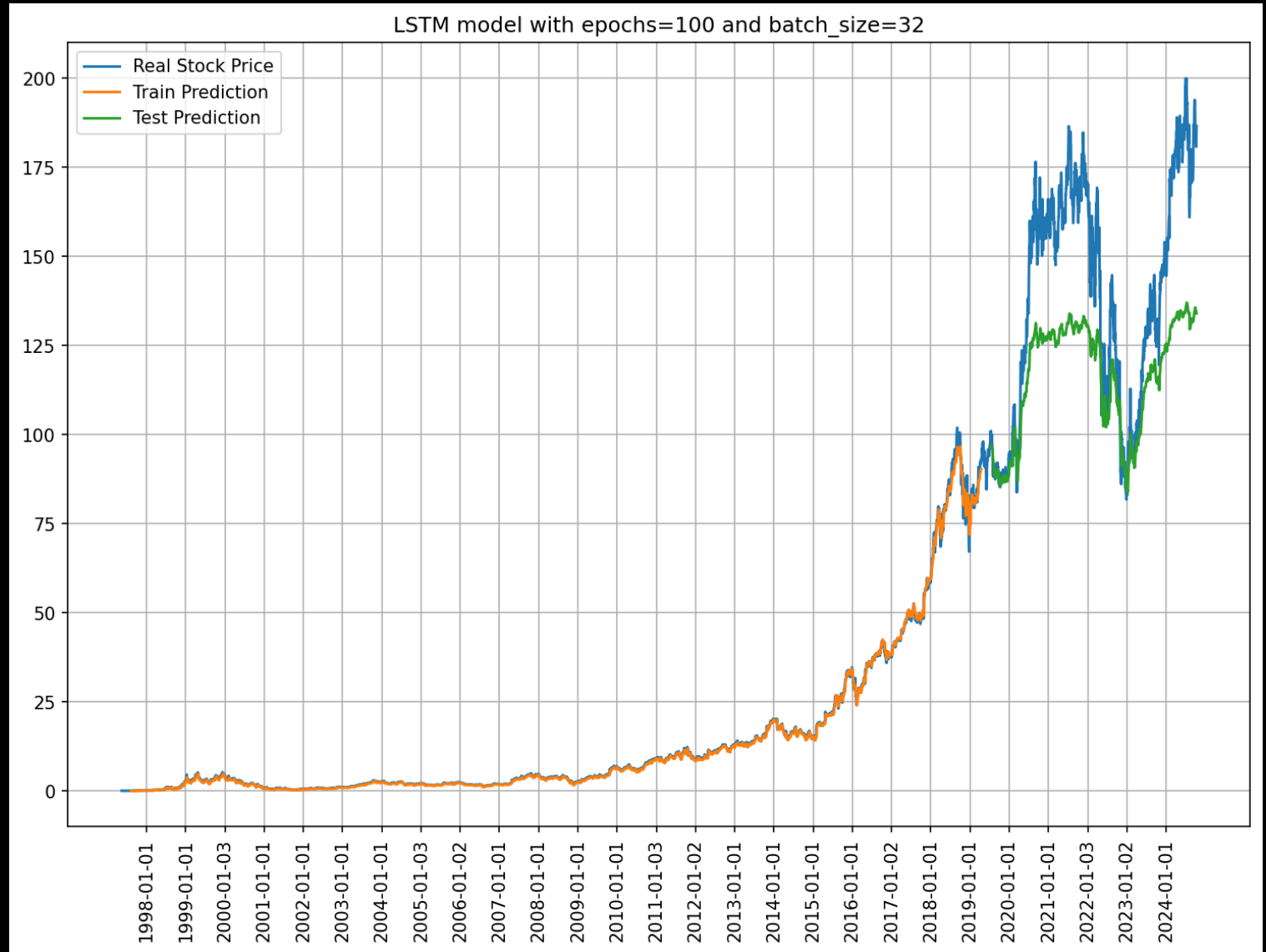
2015 - 2024



LSTM Model

epochs = 100
batch size = 32

Train rmse: 25.03
Test rmse: 116.61





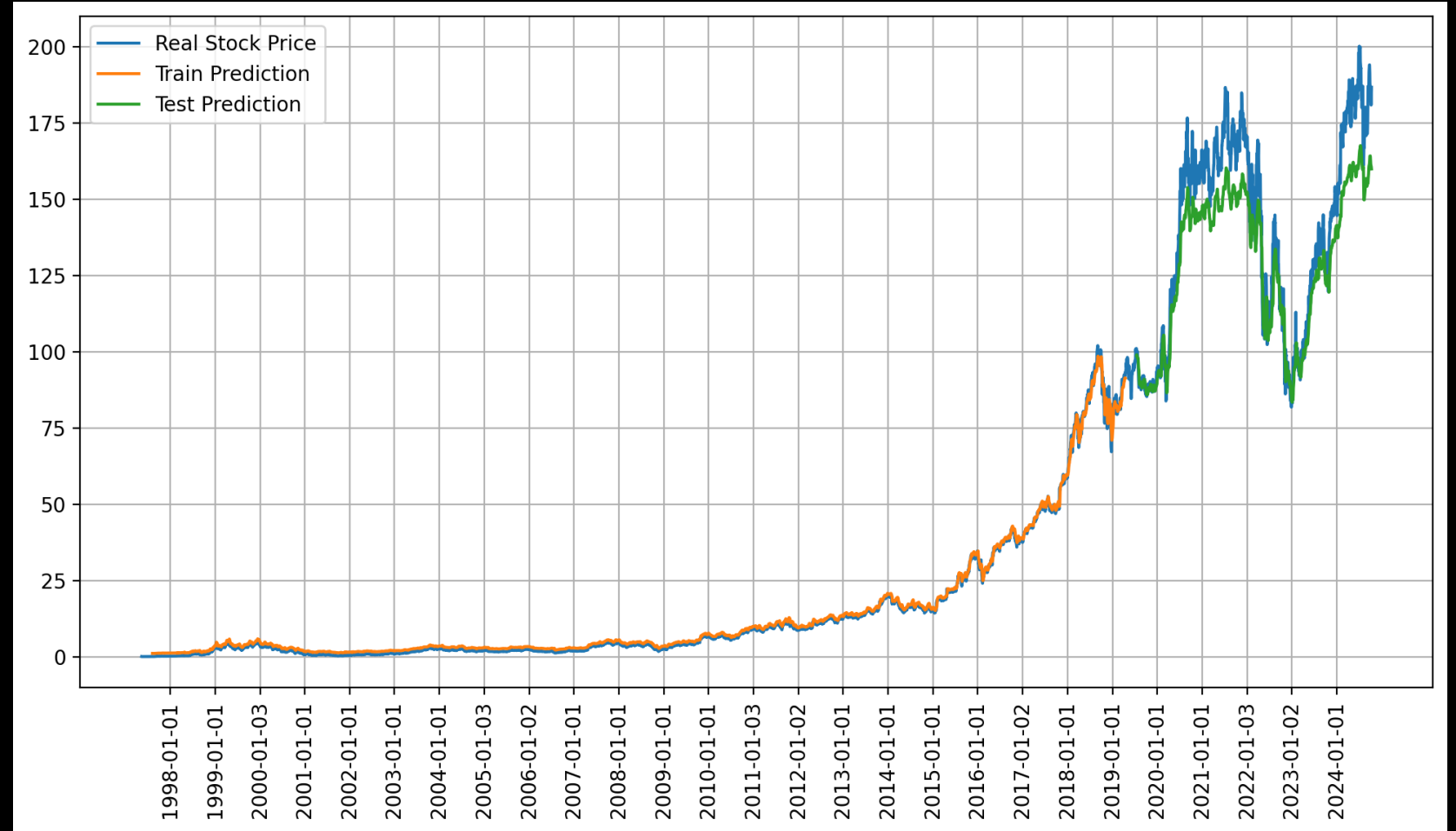
LSTM Model

epochs = 50

batch size = 32

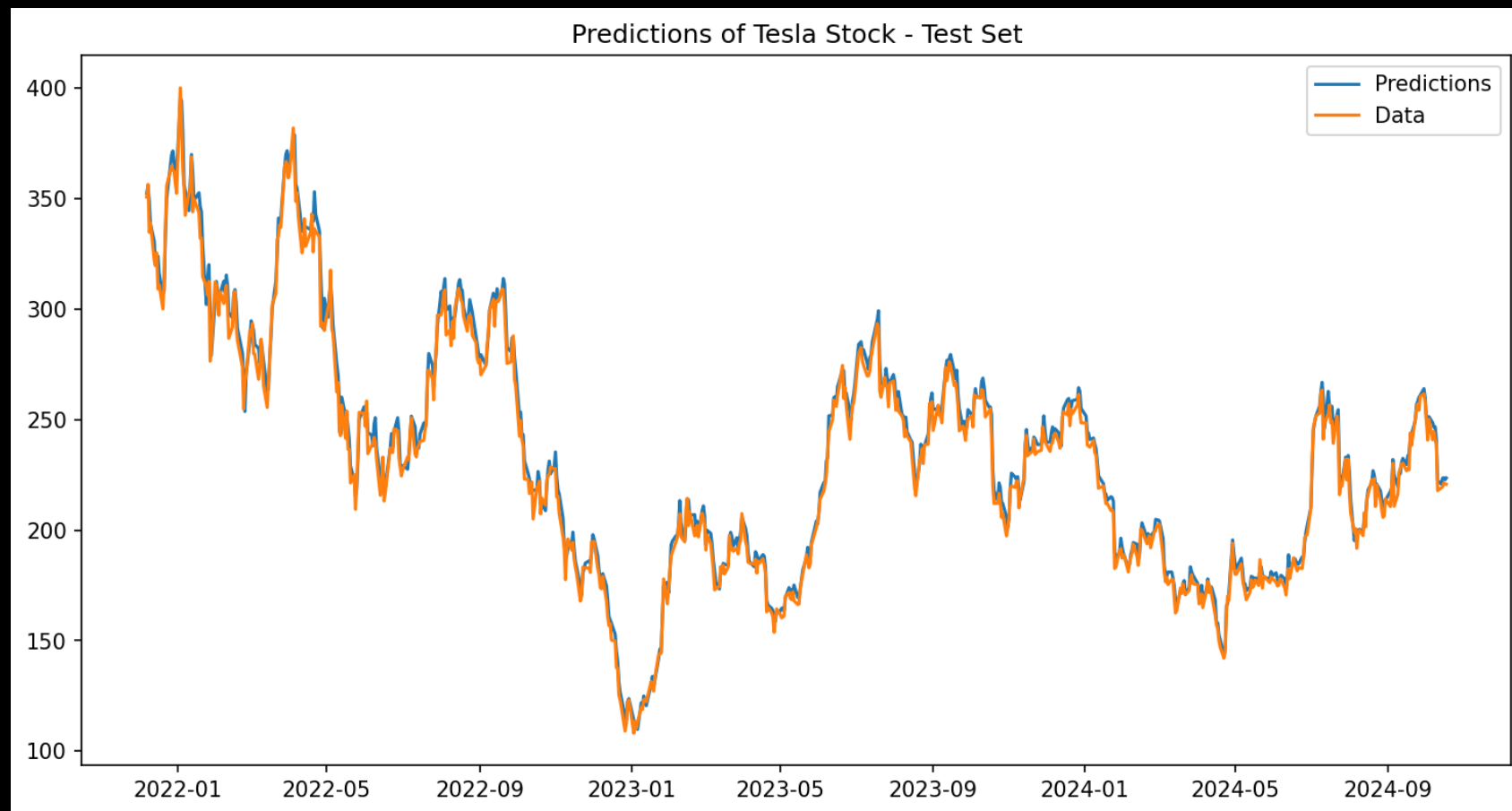
Train rmse: 25.34

Test rmse: 130.20





MLP Regression



R-squared (training set) = 0.999

R-squared (test set) = 0.989



ARIMA Model

SARIMAX Results

| | | | |
|----------------|------------------|-------------------|------------|
| Dep. Variable: | Close | No. Observations: | 3597 |
| Model: | ARIMA(1, 1, 0) | Log Likelihood | -10781.737 |
| Date: | Fri, 11 Oct 2024 | AIC | 21567.473 |
| Time: | 10:50:59 | BIC | 21579.849 |
| Sample: | 0 | HQIC | 21571.884 |
| | - 3597 | | |

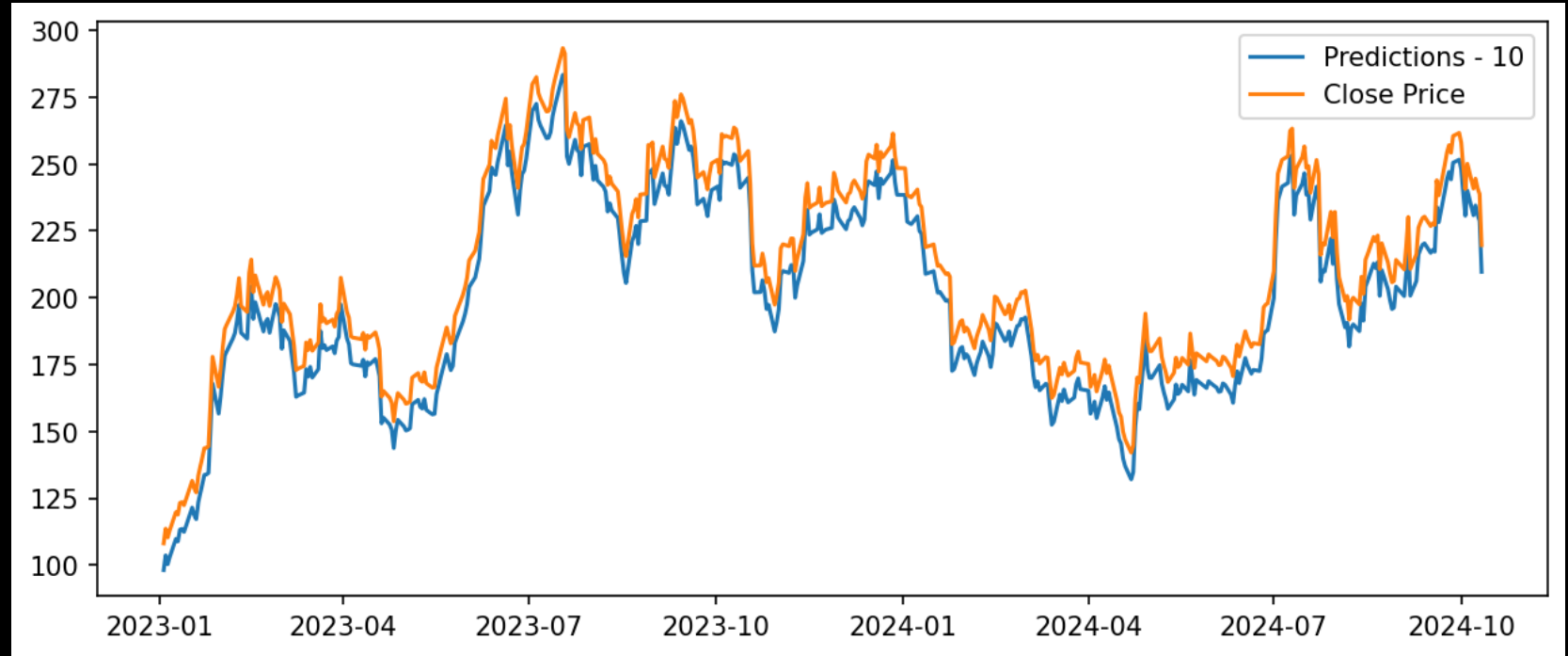
Covariance Type: opg

| | coef | std err | z | P> z | [0.025 | 0.975] |
|--------|---------|---------|---------|-------|--------|--------|
| ar.L1 | -0.0337 | 0.009 | -3.945 | 0.000 | -0.050 | -0.017 |
| sigma2 | 23.5377 | 0.183 | 128.805 | 0.000 | 23.180 | 23.896 |

| | | | |
|-------------------------|---------|-------------------|----------|
| Ljung-Box (L1) (Q): | 0.00 | Jarque-Bera (JB): | 41699.81 |
| Prob(Q): | 0.98 | Prob(JB): | 0.00 |
| Heteroskedasticity (H): | 1026.92 | Skew: | -0.20 |
| Prob(H) (two-sided): | 0.00 | Kurtosis: | 19.68 |



ARIMA Model:
Accumulative



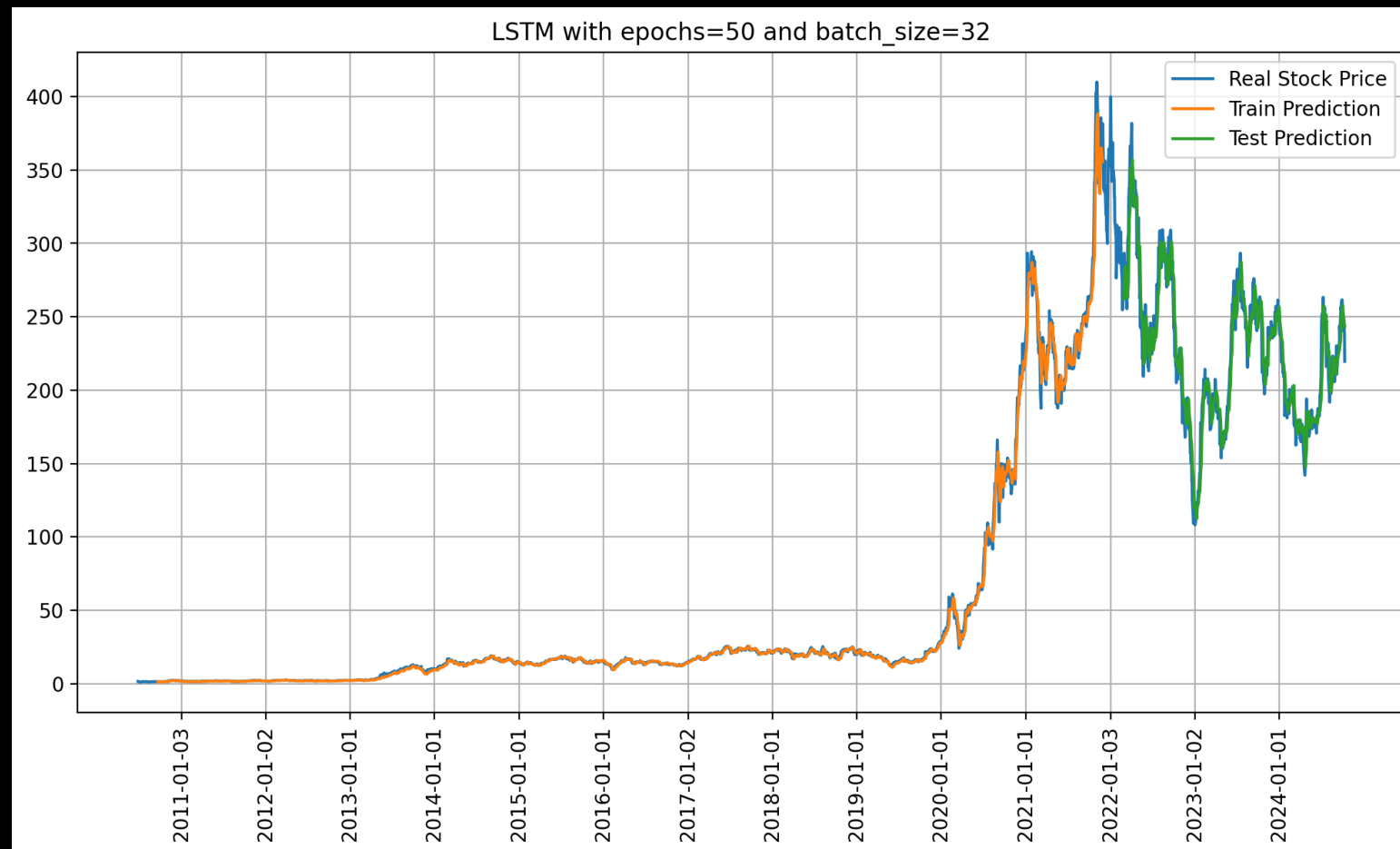
ARMA (1,1,0)



LSTM Model

epochs = 100

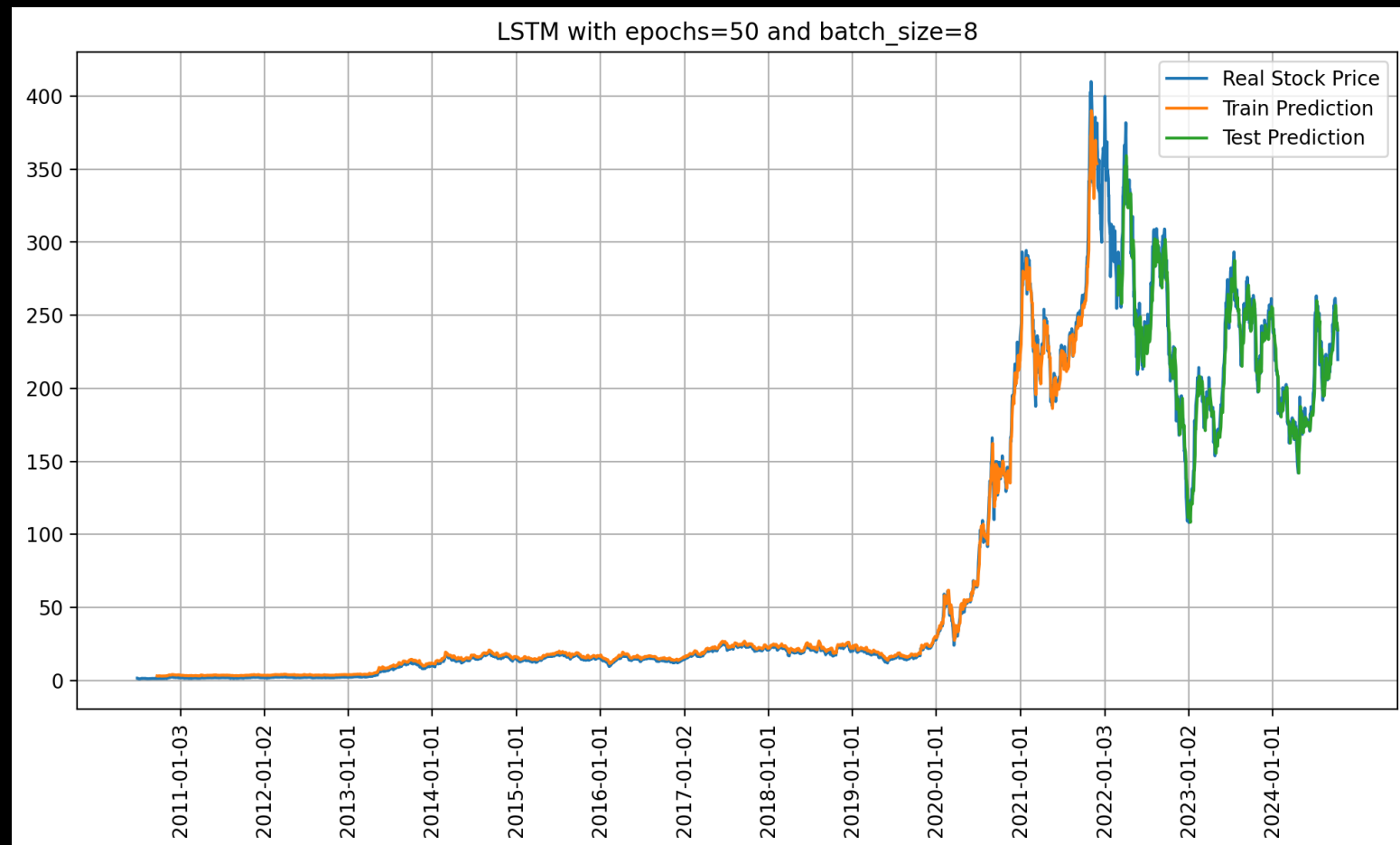
batch size = 32





LSTM Model

epochs = 50
batch size = 8



Conclusion

Supervised models – not applicable

ARIMA: better in accumulative

LSTM: the best model

Tesla vs. Amazon

