Ioanna Kontogianni

Data Bootcamp

Dr Koehler

16th December 2024

Tesla Stock Prices: A Predictive Analysis based on historical data

## Introduction

In the past 5 years, Tesla has been one of the most fast growing and trending stocks in the stock market. In my project, I aim to develop different types of predictive models in order to see how machine learning will predict the future of the stock prices of Tesla. I am able to do this through using different mathematical models and methods along side Tesla's historical data to create the new trend predictions. In my project, I use four models: my baseline which is a Naive forecasting. In addition I use a Linear Regression, Seasonal Forecasting and a Random Forest Regressor.

All these different models each offer unique insights into the complex dynamics of stock price movements. These models can be useful to not only Tesla. Using the data collected Tesla can strategically plan, consult it into their decision making and forecast their future. It can also help anyone who is interested in investing into the stock market. Through this data, an investor is able to make more informed decisions and optimize their profit to know when to buy or sell their stocks depending on where the stock predicts to trend towards.

Some of the key findings suggest that the Random Forest model provides the best predictive model followed by the Multiple Regression Model. Additionally, the Naive Forecasting model, although bad at predicting, provides for a good baseline in order to understand the other 3 models better.

Data Description

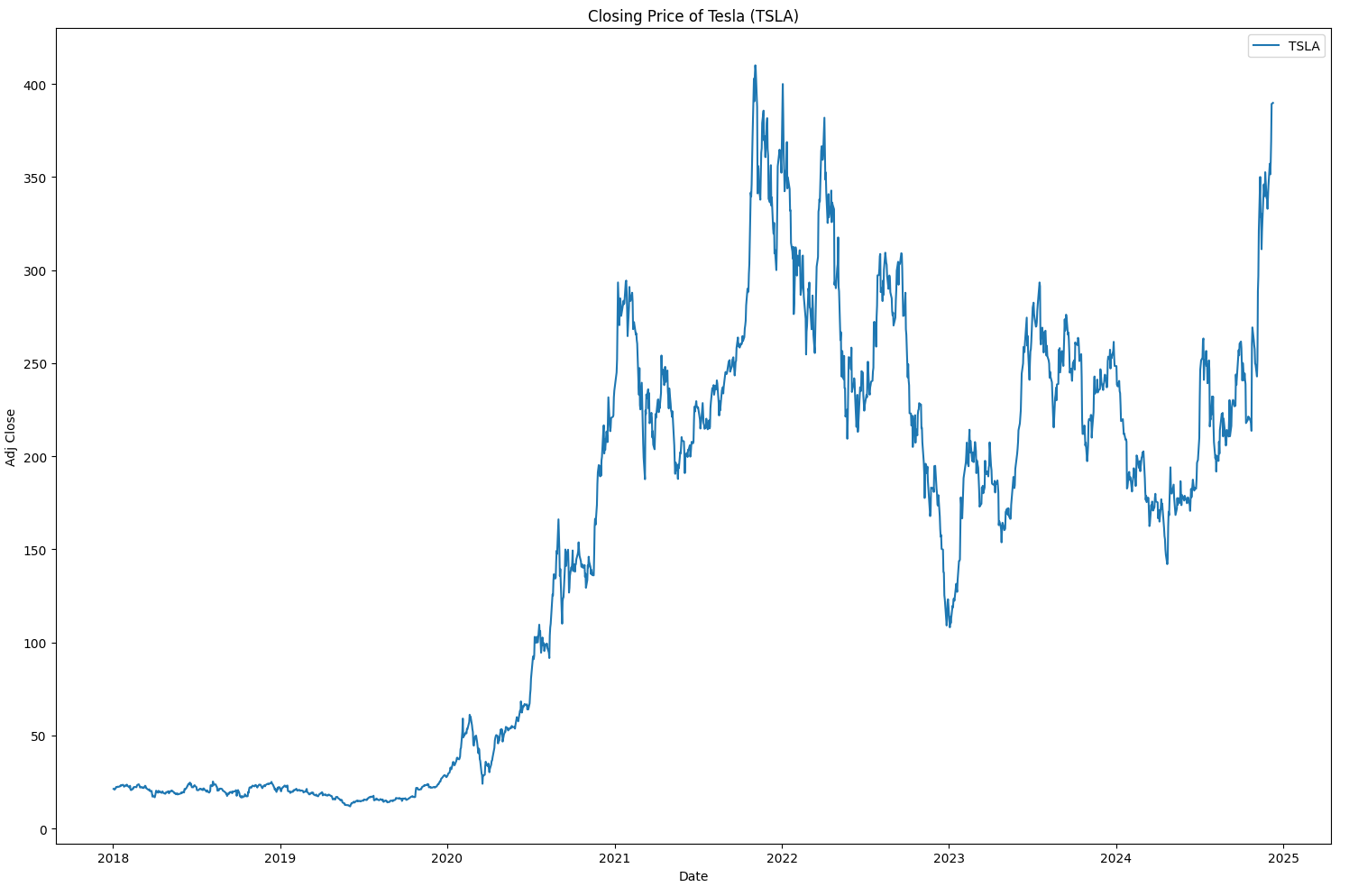
The dataset is compiled from information in the yfinance library. The company I have chosen for my models is Tesla (TSLA). The data spans eight years from January 1, 2018 to December 10th 2024.The variables that are included are:

* Close: The final price at which Tesla stock traded during a given trading day.
* Open: The price at which Tesla stock began trading when the market opened.
* High: The highest price Tesla stock reached during the trading day.
* Low: The lowest price Tesla stock reached during the trading day.
* Adj Close: The adjusted closing price, which accounts for corporate actions like stock splits or dividends.
* Volume: The number of Tesla shares traded during a given day.

This data gives the database a table of 3 columns with each variable being assigned 1746 non-null counts meaning that we have 1746 days worth of data to analyse from in those eight years.

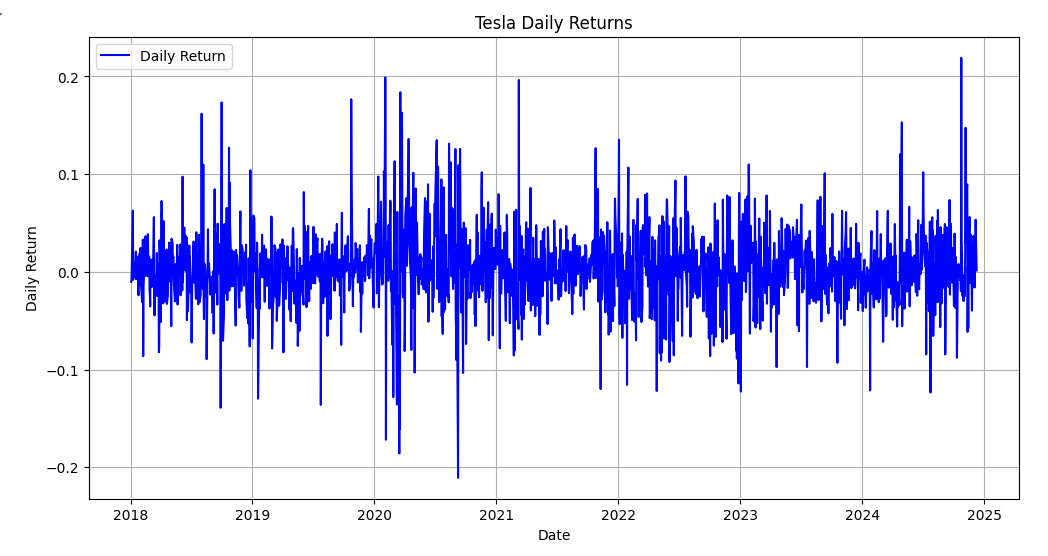
Although most of the data is price measuring variables there is also a quantity measuring variable. In my model, I look closely at the Close variable and during my description and investigation of the data, I also look closely at the Volume. The ability to use a historical analysis of this database to predict a model is that it provides valuable insights into past patterns and trends, allowing for the identification of reccuring behaviors and market dynamics to formulate.

### Historical view of Tesla's closing price



From this data we can see that during 2018 and 2019 that Tesla's stock were relatively low even though they started to trade in 2010. Using the information from Tesla's closing prices we also find that the lowest closing price was $11.93, while the highest reached $409.97 indicating significant volatility and growth in Tesla's stock price over time, even through the Covid pandemic. We also found out that the median closing price was $182.22 while the average closing price was $155.72. This means that the distribution is skewed by lower prices from before the 2020s that we can clearly see represented in the graph above. One of the most important statistics though is that the standard deviation is at $109.92, again indicating high variability in the closing prices. This means that approximately 68% of the closing prices fall within one standard deviation of the mean ($45.80 to $265.64).

### Are people who are short term investing profitable?

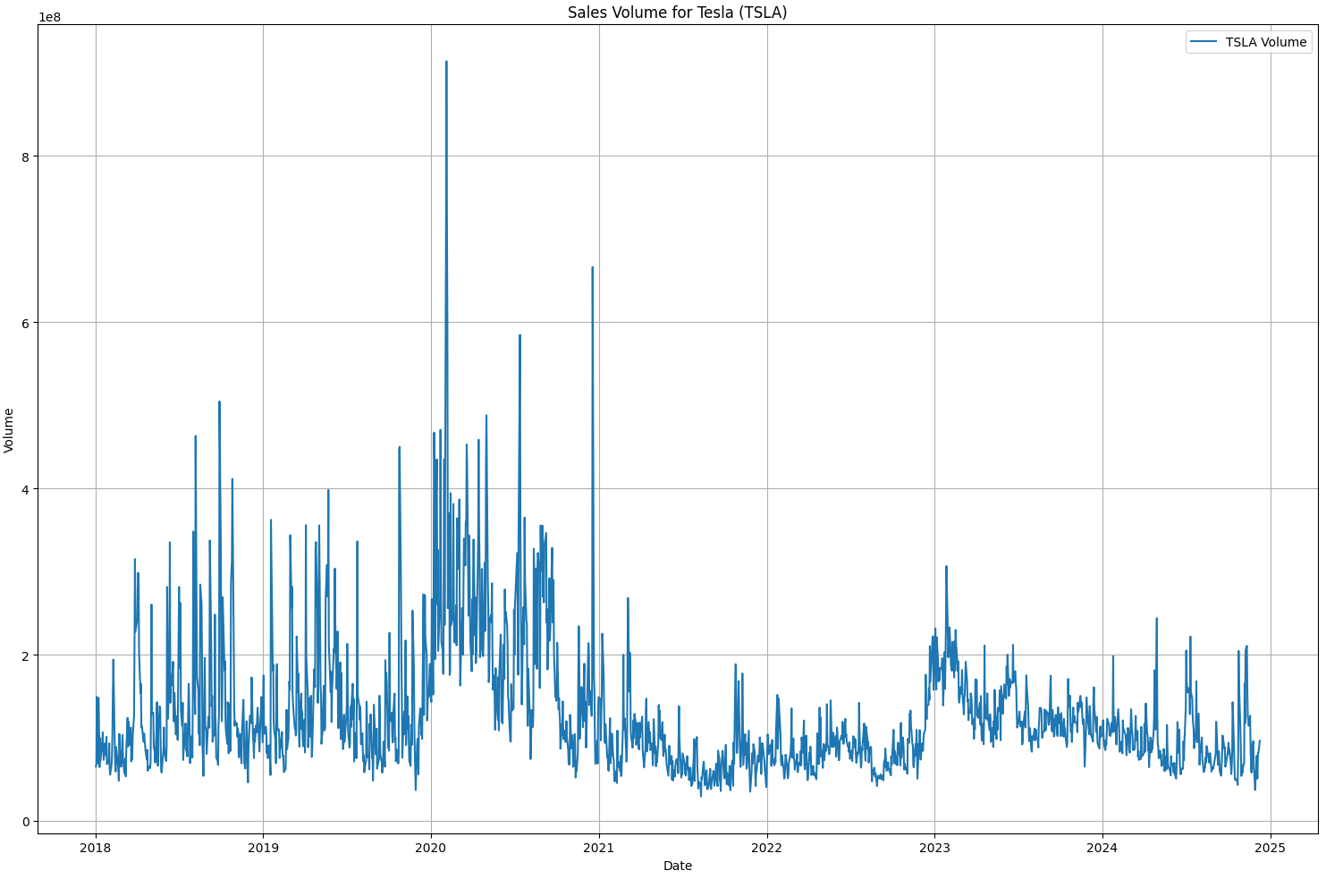


The graph shows high volatility with Sharp spikes in daily returns frequently (between +20% and -20%) indicating large daily movements. Furthermore using the statistics we see that the positive (52.15%) returns do not outweigh the negative returns (47.79%). Therefore, day to day trading for an investor in stocks is not recommended therefore what is?

### What is the most profitable duration of investing in Tesla stocks based on their previous history?

According to the statistics the percentage of positive weeks was 55.68% while the percentage of positive months was 55.42% which is considerably close. Therefore based on if returns were positive or negative burning the duration of a day, week or month, the data deems that the most profitable is investing weekly.

### Trading volume trends for Tesla



In the Sales Volume for Tesla (TSLA) graph we can see that there are several notable spikes in the volume bought especially in the start of 2020. We see a very irregular pattern between the years 2018 and 2019 and then a more calm trend after a notable spike in 2021. This is very interesting as in the closing prices we see that as time passes that price increases however we see here that the quantity of stocks bought has decreased. On the other hand it shows that sales volume has stabilized. This consistent trading volume is good as the market becomes less irregular and Tesla becomes a more trustworthy company.

## Models and Methods

For my Multiple Regression Model, I used the lagged feature in order to perform the prediction of the next day's price using the previous 5 days' prices. This was accomplished using a create\_lags() function using a 20% test split with random\_state=20 for reproducibility. This led my dataset with the dimensions (1392, 5) for training and (348, 5) for testing. This way of modelling allows for a short term price prediction however I could have used more days for increased efficiency.

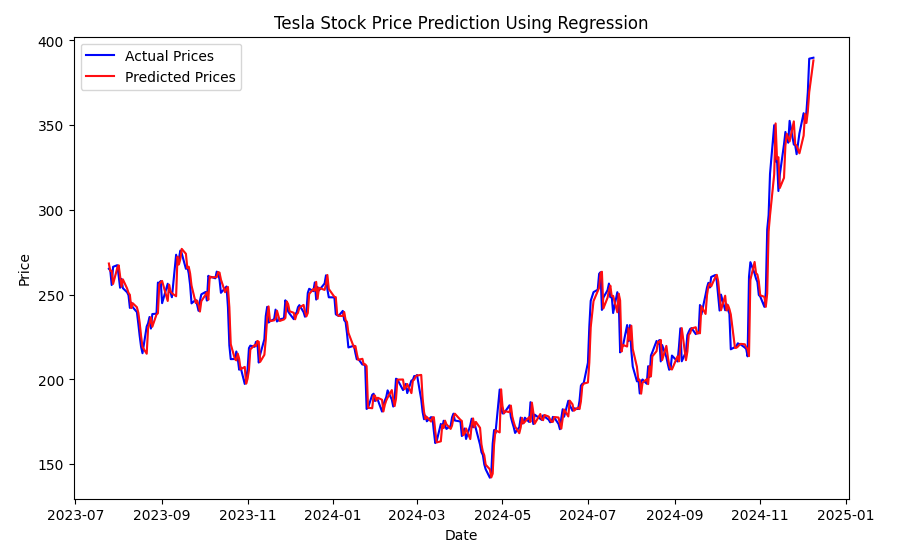
For my baseline I used a Naive Forecasting Model, which is the baseline for most time series models. One of its many advantages is that it is very simple which allows it to be very easy compared to. So ultimately, the model used the last observed closing price to project it forward into a benchmark so that ŷ(t+1) = y(t).

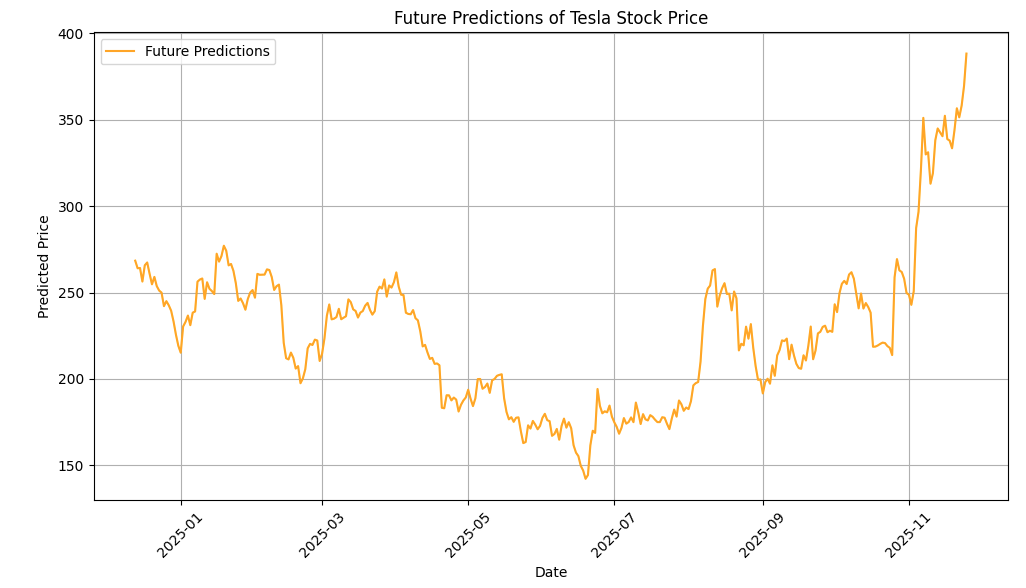
For my Seasonal Forecasting, the model captures the cyclical patterns from the previous data provided through Tesla's historical data in order to predict. In this model I was able to process this through temporal features.

Lastly for my Random Forest Model, I was able to use all the Tesla variables in order to predict the stock prices. Using n\_estimators=100 therefore 100 decision trees this model was able to train the prediction based on multiple subsets of data. This all allows me to determine features' importance and their non-linear relationship.

## Results and Interpretation

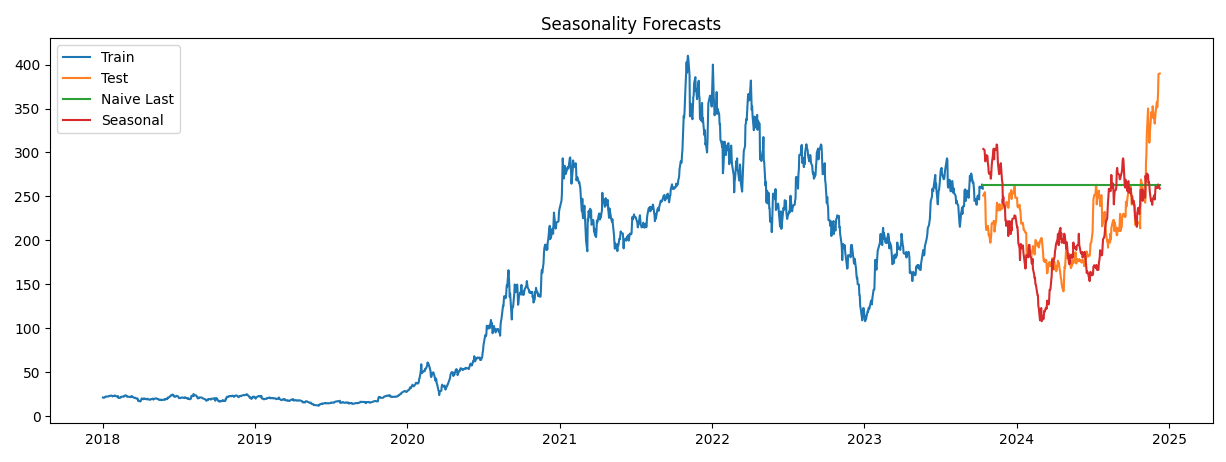
Multiple Regression Model Performance

In the model the Multiple Regression Model predicted the future with a very high score in efficiency. This can be seen in its R² score of 0.96 representing that this model can explain 96% of the variance in Tesla's stock price movements. However R² is not the evaluative tool we want to compare and contrast the rest of the models with. That would be the Mean Absolute Percentage Error (MAPE). The Multiple Regression Model’s MAPE was 2.55897% which is very small. This represents that the accuracy of the model is very high. In addition The Mean Squared Error (MSE) was 70.15, however this is quite high. For this model, the Mean Squared Error is not a very efficient way of comparing just like the R². 

Furthermore as you can see from the graph above it predicts the model very well as the lines are almost overlapping. This model is very efficient as even though there are a lot of different trends it does predict it fairly well. And that is when the graph below represents, here we see the future predictions of Tesla's stock prices.

Seasonal Forecasting

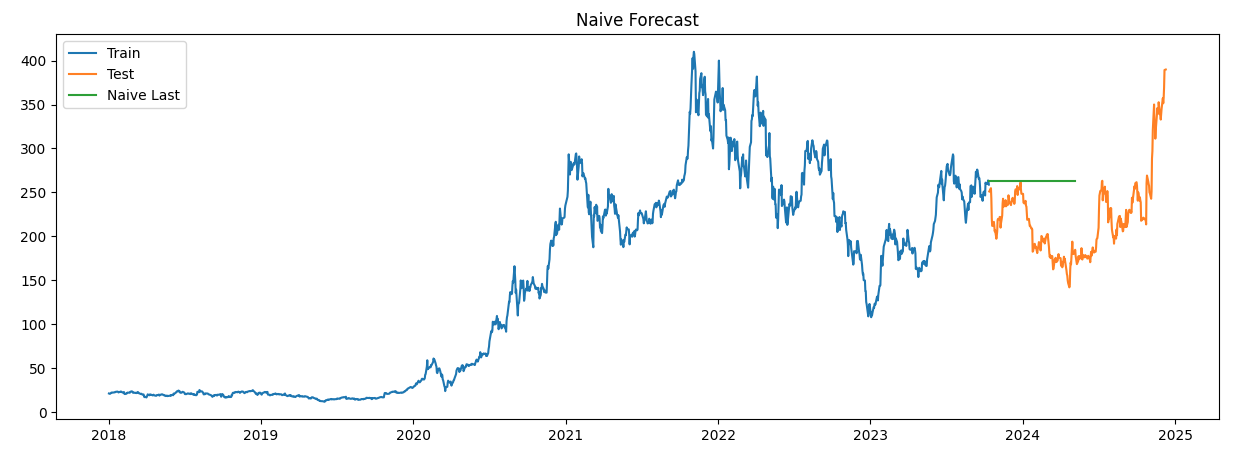
In the seasonal model we see that the predictive model did not do as good as the Multiple Regression Model even though the RMSE of the seasonal forecasting is associated with 48.36%, which is lower. The model received a MAPE of 17.411753%, telling us that there are more errors in the predictive model than in the other model.



In the graph we see that the model is not able to fully predict the magnitude of the cyclical patterns and trends. The visualization shows us that this model has moderate predictive capabilities and has potential however is unable to deal with high volatility in the data. However compared to the baseline, this model represents the data better.

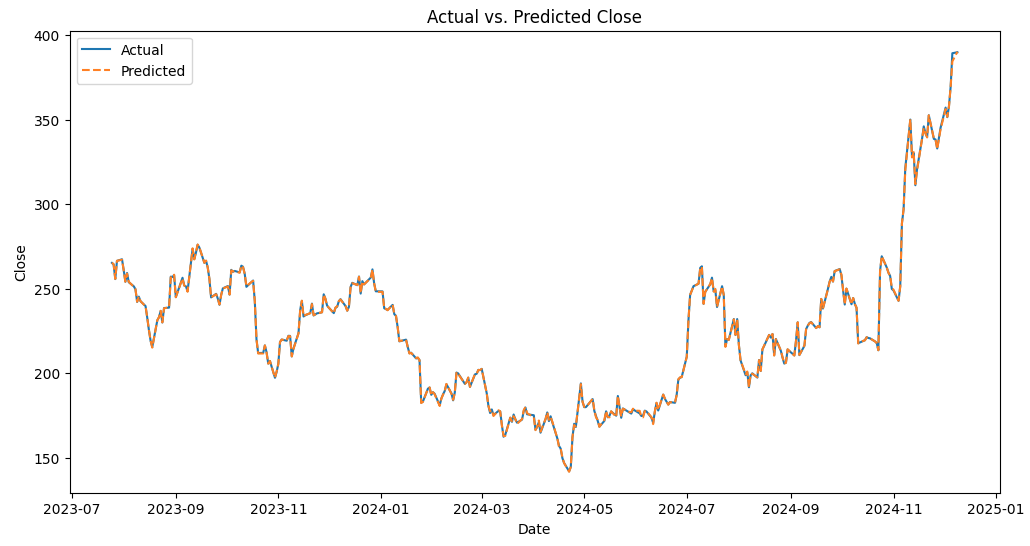
Naive Forecast Performance

The naive forecasting model acts as a baseline therefore is not the best model of fit this can be seen as it keeps a constant predict line at around 250. Due to this its RMSE is 62.52 and unsurprisingly its MAPE is 27.24% which is relatively high. This further shows us that this baseline model does not have good predictive capabilities and that there are significant limitations. This is further represented by the graph below as the straight line shows the Naive Forecast does not account at all for any market change and is just a horizontal line.



Random Forest Model Performance

The Random Forest Model in this model shows great accuracy and efficiency in predicting Tesla’s Stocks. This is translated very obviously in the data as the Random Forest Models performance is very good. RMSE is 0.414 while Mean Absolute Percentage Error is equal to 0.109140%. Both of these values show the reliability of this model, however nothing like the graph below that shows that this model is able to predict Tesla’s stock close price to a great extent even in high volatility.



Comparative Analysis

When comparing all four models we can only really compare them with only one metric which is the MAPE. Using the Mean Absolute Percentage we see that the Random Forest model is the best followed by the Multiple Regression model, the Seasonal Forecast and then lastly the Naive Forecasting. It is very interesting to see that even for the RMSE the order is still the same however for this context I will be only evaluating the model based on this because MAPE is more scale dependent and for this contect more reasonable to. Using the MAPE I am able to communicate the risk to an investor who can easily interpret as it expresses error as a percentage which will allow for better interpretation. Furthermore RSME is more sensitive to outliers and is harder to interpret. The Random Forest model was the best approach in predicting data and so was the Multiple Regression model as both used very high accuracy to predict the closing prices. Although better than the baseline, the Seasonal Forecasting did not quite predict the closing price as accurately as it should have. Finally the baseline model, The Naive Forecasting demonstrates that predicting stocks with very simple models is not possible especially in such dynamic markets.

Conclusion and Next Steps

All of these methods of predicting models have allowed for great exploration of the analysis of Tesla’s future stock prices. While the MAPE and RMSE scores told us the predictive capabilities of each model, the visualization of the models is what really represented the data the best, especially since my thesis is aimed at trying to help investors. The Random Forest model showed us an outstanding predictive model followed by the Multiple Regression Model that also performed very well but not quite capturing market dynamics. Additionally, the Naive Forecasting model represented a good simple baseline model which helped us compare.

In order to improve I would love to add more models to this analysis or even better improve on these models such as using additive and multiplicative models to my Seasonal model. In addition, for this model, I only used seven years, however for a next step, I could look back at all the data. Increasing my dataset will increase my accuracy on training and testing the predictive models. Furthermore, I could use more tickets such as other companies to see if there is any correlation between competitors in stock prices. Lastly, I could use the Adj Close price to better the model for investors to use.