
# Tesla Stock Price Forecasting

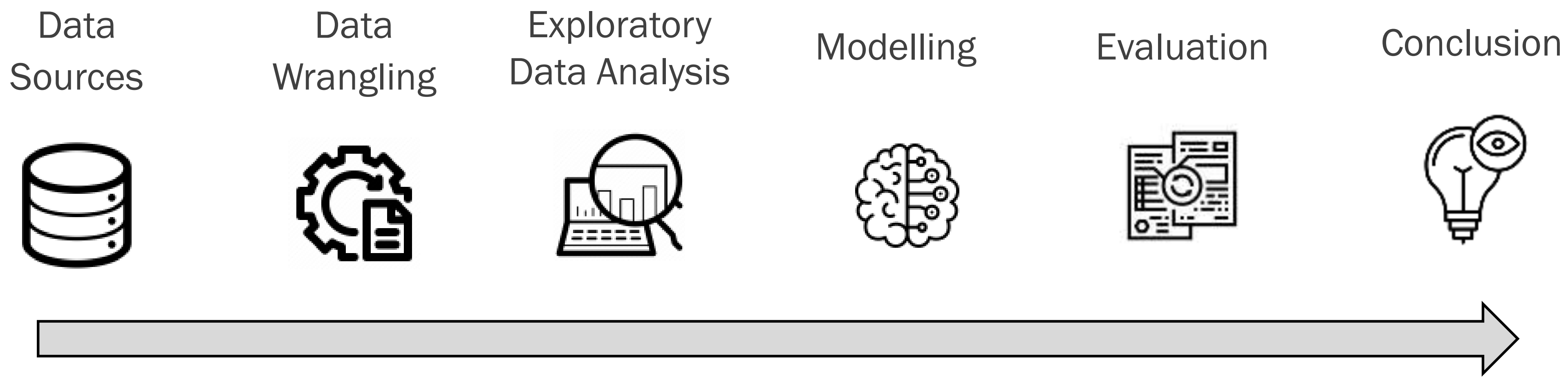
CAPSTONE PROJECT

DIEGO MEZA DELGADO

# Context

- Finance and trading companies need accurate models to predict stock prices to make informed investment decisions.
- Tesla's stock price is highly volatile, making it difficult to predict with traditional methods.
- The stock market is influenced by a range of factors: economic indicators, correlated stocks, and other market dynamics.
- The goal of this project is to leverage historical data (Tesla and correlated stocks) and economic indicators to forecast Tesla's stock price.

# Approach


Data Sources
Data Wrangling
Exploratory Data Analysis
Modelling
Evaluation
Conclusion

# The data

**Stock Data (Yahoo Finance API - yfinance):**

- Tesla Stock Price: Primary target for forecasting.
- Correlated Stocks: Nio, Rivian, Lucid, Ford, and GM were chosen to account for competitors and market influences.

**Economic Indicators:**

- Interest Rate & Inflation (CPI): From FRED, reflecting macroeconomic conditions impacting stock prices.
- Consumer Confidence: From Investing.com, to measure market sentiment and consumer behavior.
- Oil Price: From Macrotrends, as it affects production costs and profitability in the automotive sector.

# Data Wrangling

### Index Alignment

Synchronized all time series data (Tesla, correlated stocks, and economic indicators) to ensure consistency in the time period.

### Interpolation

Converted monthly economic data (e.g., Interest Rate, CPI) into daily values to match the stock price frequency.

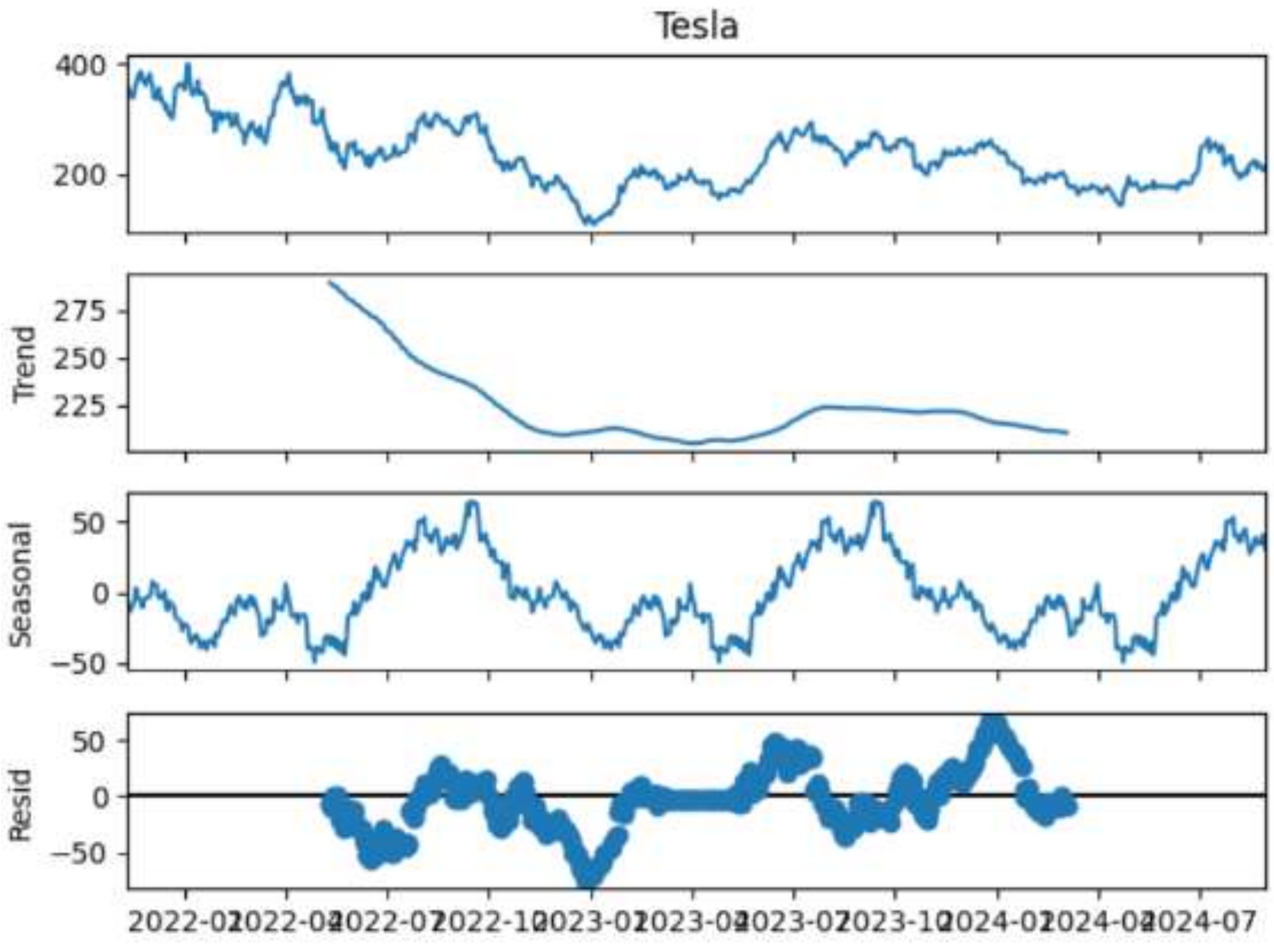
### Data Slicing

Adjusted all data sources to have the same start and end dates, ensuring a unified dataset for modeling.

# Exploratory Data Analysis

## Time Series Decomposition

Plotting the components helps in understanding the underlying patterns and periodic behaviors in the data: Trend (long-term movement), Seasonal (regular, repeating patterns), and Residual (random noise).



# Exploratory Data Analysis

## Stationarity Test

Augmented Dickey-Fuller Test: The ADF Test was applied to assess whether each time series is stationary. A stationary time series has constant mean and variance over time, which is a crucial assumption for many time series models.

```
Tesla
ADF Statistic: -2.9710149207598815
p-value: 0.03769134077959895
Critical Values: {'1%': -3.4398077121659765, '5%': -2.865713608066101, '10%': -2.5689925469026402}
Conclusion: Reject null hypothesis - Time series is stationary
----------------------------------------------------------------------------------------------------
Nio
ADF Statistic: -3.14920272439644
p-value: 0.023119829824608057
Critical Values: {'1%': -3.439918423003054, '5%': -2.865762386436236, '10%': -2.5690185346241785}
Conclusion: Reject null hypothesis - Time series is stationary
----------------------------------------------------------------------------------------------------
```

# Exploratory Data Analysis

## Correlation Test

Correlation analysis was performed to evaluate the strength and direction of the relationship between Tesla's stock price and other time series (correlated stocks and economic indicators). This helps in identifying which variables have a significant linear relationship with Tesla's stock price.

```
Pearson Correlation Coefficient: 1.0
Pearson Correlation Coefficient: 0.7705541831309725
Pearson Correlation Coefficient: 0.7451335448301191
Pearson Correlation Coefficient: 0.7550701196501474
Pearson Correlation Coefficient: 0.6683361725440896
Pearson Correlation Coefficient: 0.44429820194863673
```

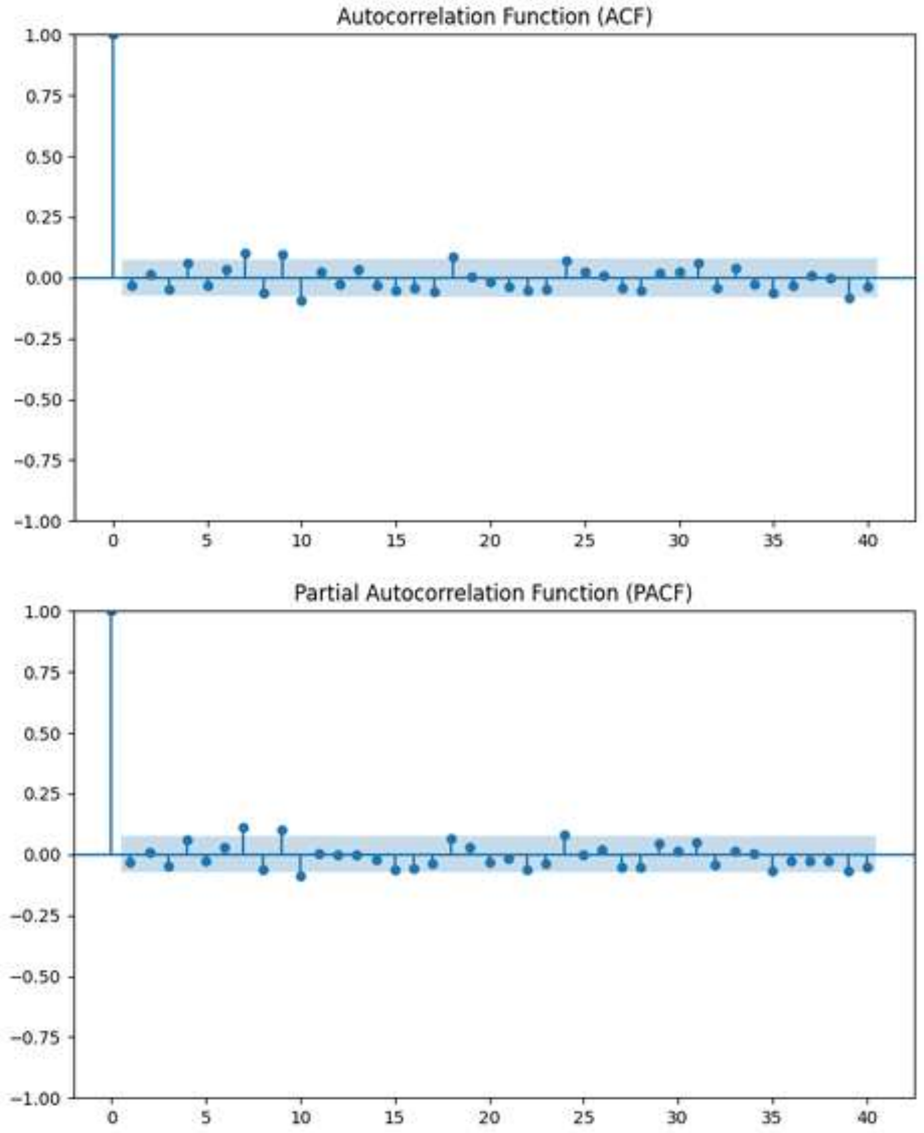
```
Pearson Correlation Coefficient: -0.7538193287726164
Pearson Correlation Coefficient: 0.4533791266639633
Pearson Correlation Coefficient: 0.29439013888901056
Pearson Correlation Coefficient: 0.23805135167273392
```

| Scale of correlation coefficient | Value |
|---|---|
| $0 < r \leq 0.19$ | Very Low Correlation |
| $0.2 \leq r \leq 0.39$ | Low Correlation |
| $0.4 \leq r \leq 0.59$ | Moderate Correlation |
| $0.6 \leq r \leq 0.79$ | High Correlation |
| $0.8 \leq r \leq 1.0$ | Very High Correlation |

# Exploratory Data Analysis

## ACF and PACF

The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots were used to visualize the correlation of a time series with its own past values. The ACF shows the correlation of the series with lags of itself, while the PACF shows the correlation after removing the effect of intermediate lags.



# Preprocessing

| | |
|---|---|
| **Missing Values** | Checked for missing values in the dataset, and none were found. |
| **Feature Engineering** | Date Related Features: Month, day, weekday and year.<br>Lag Features: Rolling MEAN (1, 4, 30) and rolling STD (4, 30). |
| **Set Time Series Frequency** | Data frequency was set to daily to ensure compatibility with ARIMA modeling. |
| **Train-Test Split** | Train Set: First 80% and Test Set: Last 20%. |

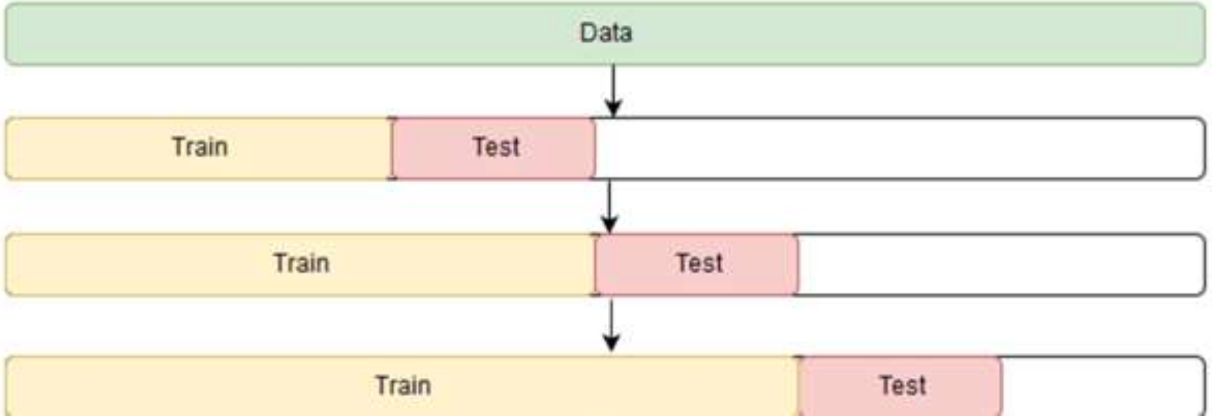
# Modelling

## Time Series Cross Validation

Fold 1: Training set of 205 data points, validation set of 205 data points.

Fold 2: Training set of 410 data points, validation set of 205 data points.

Fold 3: Training set of 615 data points, validation set of 205 data points.



# Modelling

### Statistical Models

ARIMA (0,1,0)
ARIMA (1,1,0)
ARIMA (0,1,1)
ARIMA (1,1,1)

ARIMAX (0,1,0)
ARIMAX (1,1,0)
ARIMAX (0,1,1)
ARIMAX (1,1,1)

### ML Models

Linear Regression
Lasso Regression
Ridge Regression
XGB Regressor

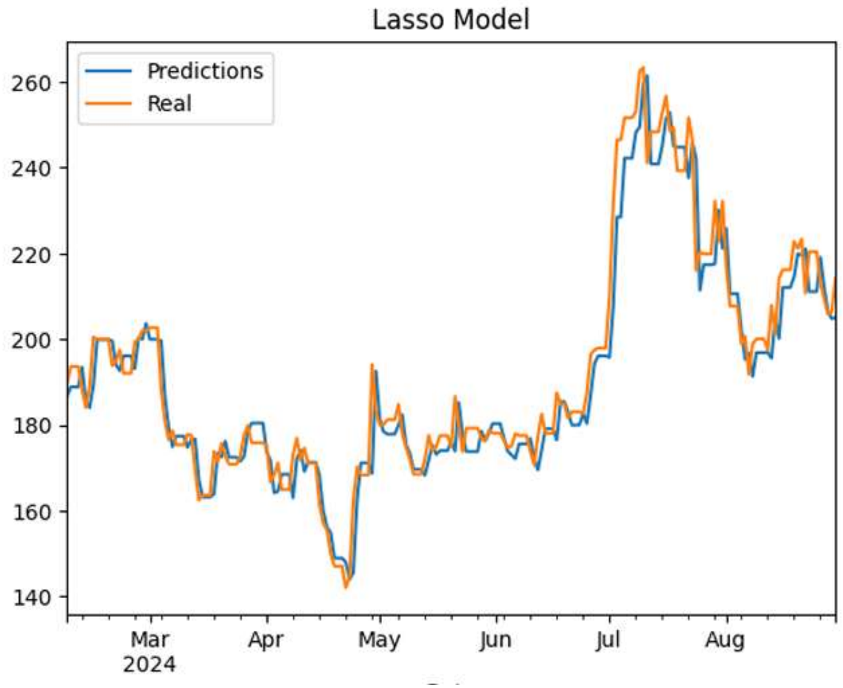
| | MAE | RMSE | AIC | BIC |
|---|---|---|---|---|
| **ARIMA(0,1,0)** | 54.93 | 63.17 | 2967.94 | 2971.85 |
| **ARIMA(1,1,0)** | 54.76 | 62.99 | 2969.13 | 2976.96 |
| **ARIMA(0,1,1)** | 54.75 | 62.98 | 2969.13 | 2976.96 |
| **ARIMA(1,1,1)** | 54.74 | 62.93 | 2970.74 | 2982.49 |
| **ARIMAX(0,1,0)** | 148.99 | 162.16 | 2717.96 | 2968.67 |
| **ARIMAX(1,1,0)** | 151.78 | 164.99 | 2717.50 | 2972.12 |
| **ARIMAX(0,1,1)** | 151.87 | 165.18 | 2717.48 | 2972.10 |
| **ARIMAX(1,1,1)** | 151.56 | 164.85 | 2718.42 | 2976.96 |
| **Linear Regression** | 193.39 | 212.95 | NaN | NaN |
| **Ridge** | 23.16 | 28.31 | NaN | NaN |
| **Lasso** | 6.41 | 8.39 | NaN | NaN |
| **XGB** | 26.44 | 34.06 | NaN | NaN |

# Evaluation

Hybrid Model = ARIMA (1,1,1) + Lasso

Hybrid Model was proved on the test set, searching for a weight for ARIMA and Lasso. The optimal weight is 100% for Lasso forecasts. The metrics of the final model are:

```
RMSE Final Model: 6.53
MAE Final Model: 4.75
```



# Conclusion

In this project, we explored various methods for forecasting Tesla's stock price using a combination of statistical models (ARIMA and ARIMAX), machine learning models (Linear, Lasso, Ridge), and an ensemble method (XGB Regressor). After thorough evaluation using time series cross-validation, Lasso Regression emerged as the most effective model, outperforming both the ARIMA-based models and the hybrid approach. Despite experimenting with hybrid models, the Lasso model alone provided the most accurate predictions, as confirmed by its optimized performance and alignment with the test set. This result highlights the strength of machine learning approaches, particularly Lasso Regression, for forecasting complex time series data in financial markets. Future work could explore additional features, alternative machine learning techniques, or even incorporating external sentiment analysis to further refine predictions.

- Incorporate Alternative Data Sources like social media for sentiment analysis.
- Explore Advanced Machine Learning Models like Long Short-Term Memory Networks.
- Refine feature selection.
- Expand Training Data.
- Regular Model Retraining.
- Risk Management integrating the model into broader risk management strategies.

# Final Recommendations

# Thanks