Stock Price Prediction Pipeline Evaluation

Stock: TSLA

Fetch ID: fetch_20250617_093553

Model ID: model_tsla_20250815_170544

Variants: with_outliers, without_outliers

Date: August 15, 2025

Author: Diego Lozano

Project Overview

This report examines a machine learning pipeline designed to predict stock price movements, adaptable to any stock symbol. The pipeline leverages historical data and key market features to assess predictive performance across various conditions. Visualizations illustrate the accuracy and limitations of the predictions, particularly during volatile periods.

Date: August 15, 2025

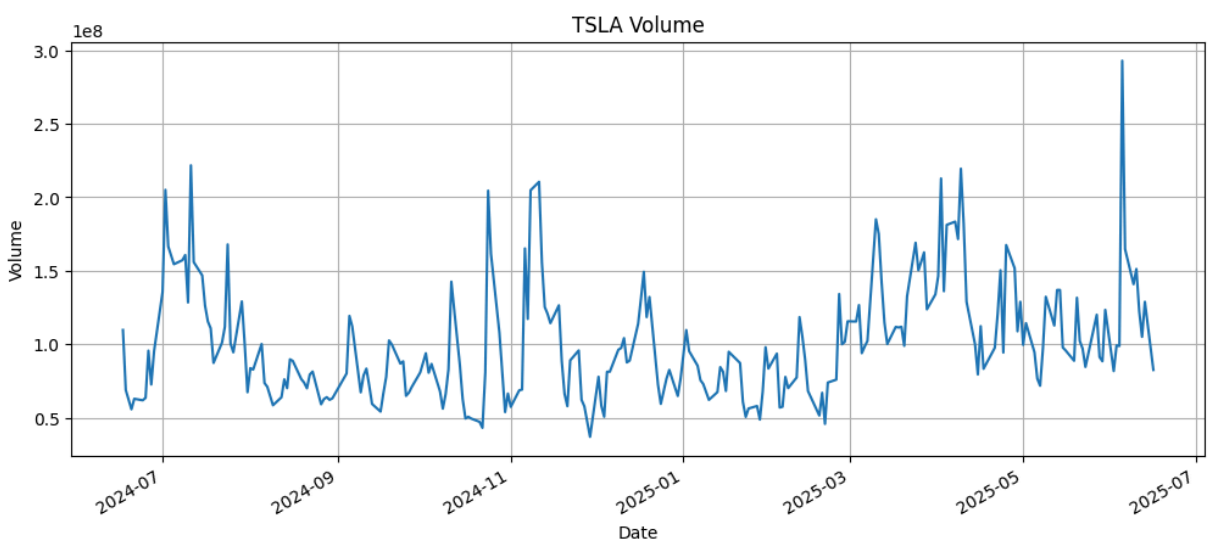
Stage: Inspect Raw Data

Objectives:

- Verify raw stock data integrity (row count, date range, data types).
- Identify gaps in trading days, missing values, and anomalies.
- Visualize closing price and trading volume to detect trends or outliers.

Visualization

Visualizations of closing price and trading volume help identify trends and anomalies visually, complementing the statistical analysis.



Quantitative Summary

Data

- **Row Count**: 250
- **Date Range**: 2024-06-17 to 2025-06-16
- **Missing Values**: 0
- **Anomalies**: 1

Anomalies

	date	column	value
1	2025-06-05	volume	292,818,655.00

Statistics

	open	high	low	close	volume
count	250	250	250	250	250
mean	290.45	297.72	282.84	290.46	100,814,958
std	72.62	74.00	70.21	71.99	39,521,561
min	177.92	183.95	177.00	181.57	37,167,621
25%	230.44	237.30	225.10	231.46	71,340,436
50%	263.05	274.11	257.10	265.41	93,858,026
75%	345.07	351.22	335.92	344.16	120,581,185
max	475.90	488.54	457.51	479.86	292,818,655

Date: August 15, 2025

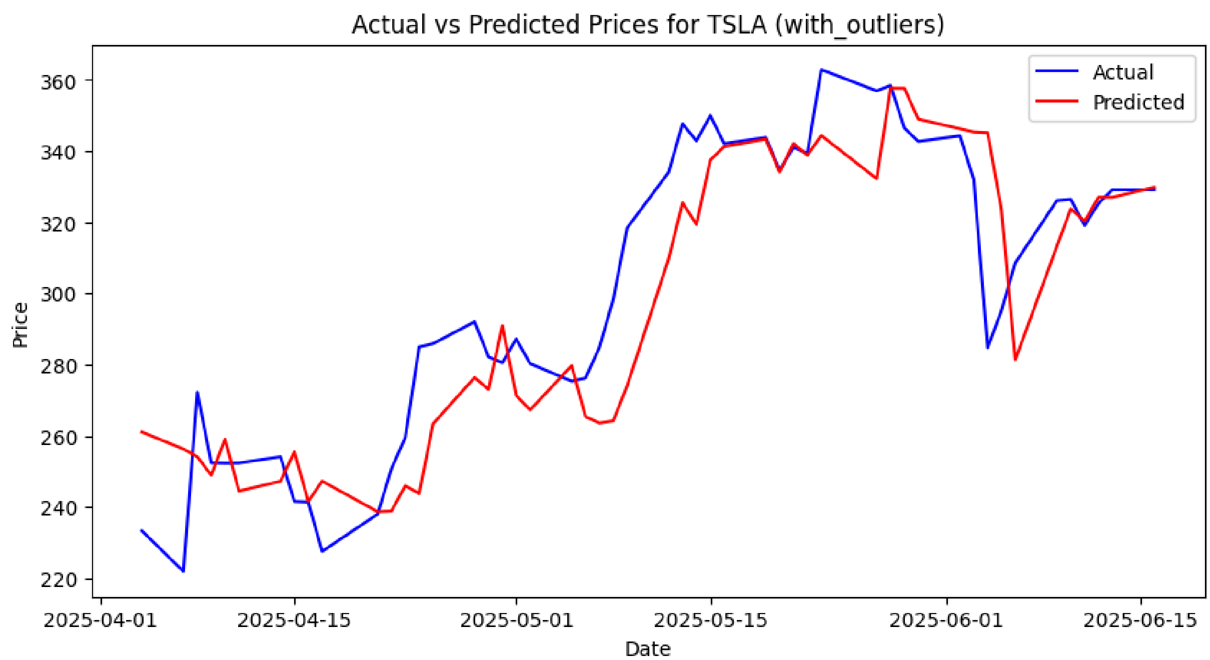
Stage: Model Performance Analysis

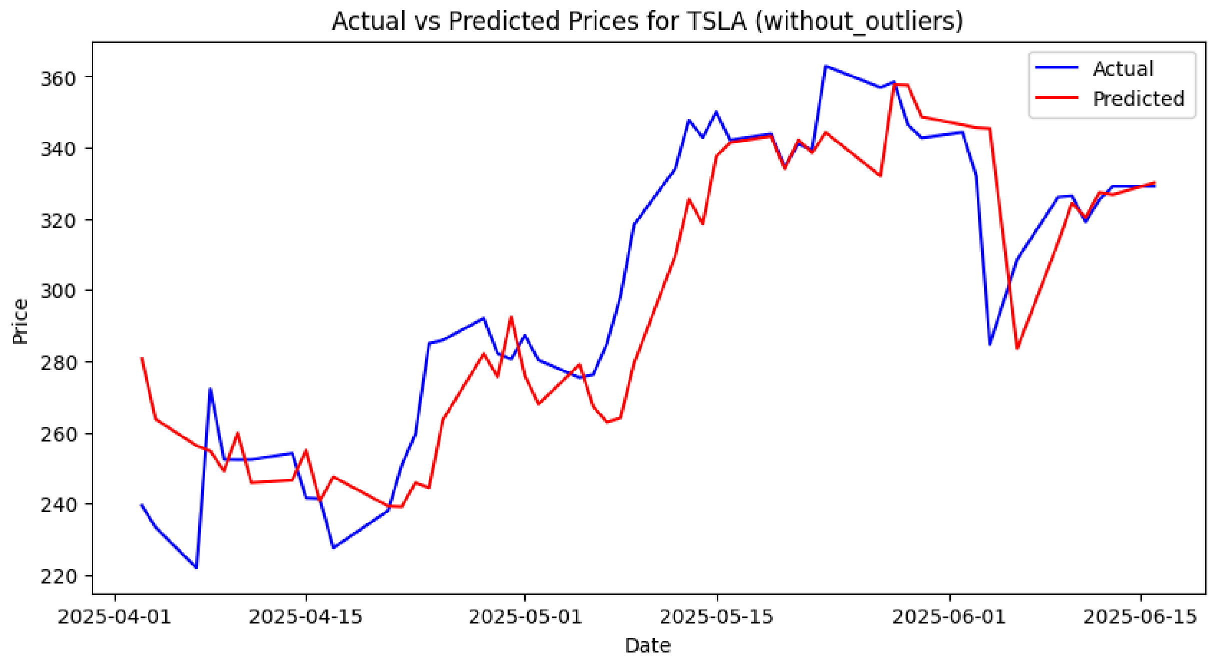
Objectives:

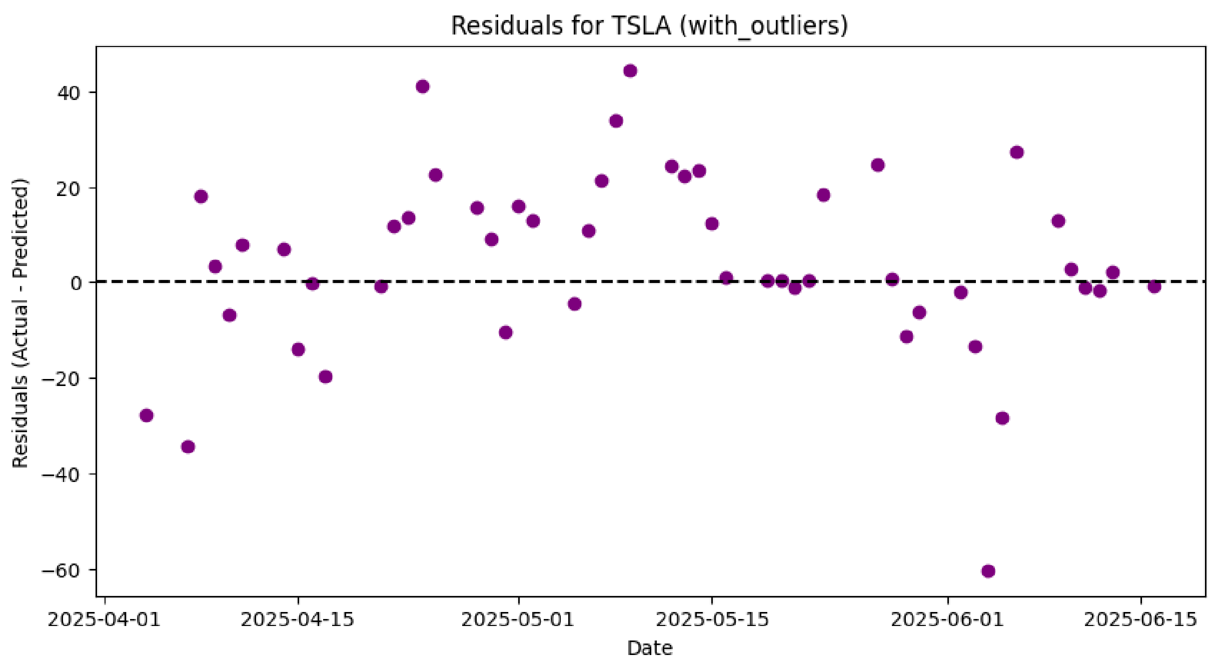
- Visualize predictions and residuals to assess model accuracy.
- Evaluate linear regression model performance (with/without outliers) using RMSE, MAE, and R^2.
- Analyze the impact of outliers on model performance.

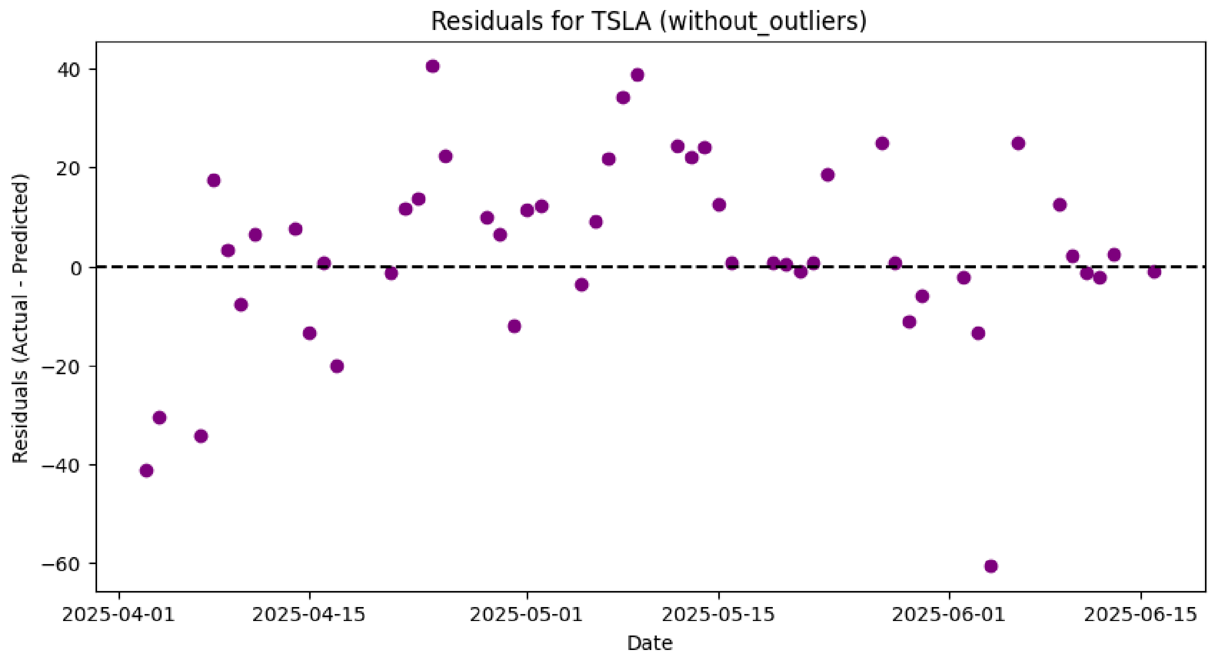
Visualization

Prediction vs. actual plots compare predicted and actual values, showing alignment with occasional deviations during volatile periods. Residual plots illustrate prediction error distribution, highlighting potential over- or underprediction, especially during market shifts. These visualizations are available for review.




Actual vs Predicted Prices for TSLA (without_outliers)
Actual
Predicted
Price
220
240
260
280
300
320
340
360
2025-04-01
2025-04-15
2025-05-01
2025-05-15
2025-06-01
2025-06-15
Date


Residuals for TSLA (with_outliers)
Residuals (Actual - Predicted)
−60
−40
−20
0
20
40
2025-04-01
2025-04-15
2025-05-01
2025-05-15
2025-06-01
2025-06-15
Date



Model performance and outliers impact

- **Model with Outliers**:
    - RMSE: 19.30
    - MAE: 14.16
    - R^2: 0.77
- **Model without Outliers**:
    - RMSE: 19.39
    - MAE: 14.07
    - R^2: 0.78
- **Outlier Impact**:
    - Date 2025-06-05: Error with outliers: 28.45, Error without outliers: 28.37