

Date: August 01, 2025

Stage: Model Performance Analysis

Objectives

- Evaluate the performance of linear regression models (with and without outliers) using metrics such as RMSE, MAE, and R^2 .
- Visualize predictions and residuals to assess model accuracy.
- Analyze the impact of outliers (e.g., June 5, 2025) on model performance.

Model Evaluation

Model with outliers:

RMSE: 19.54

MAE: 14.47

R^2 : 0.77

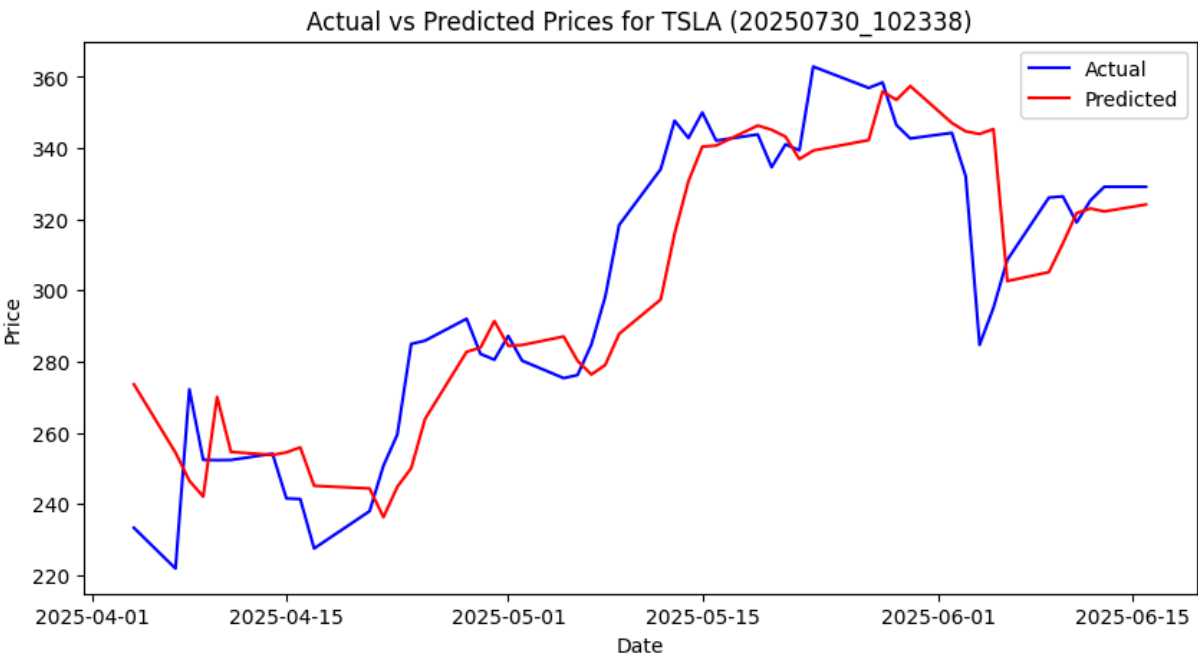
Model without outliers:

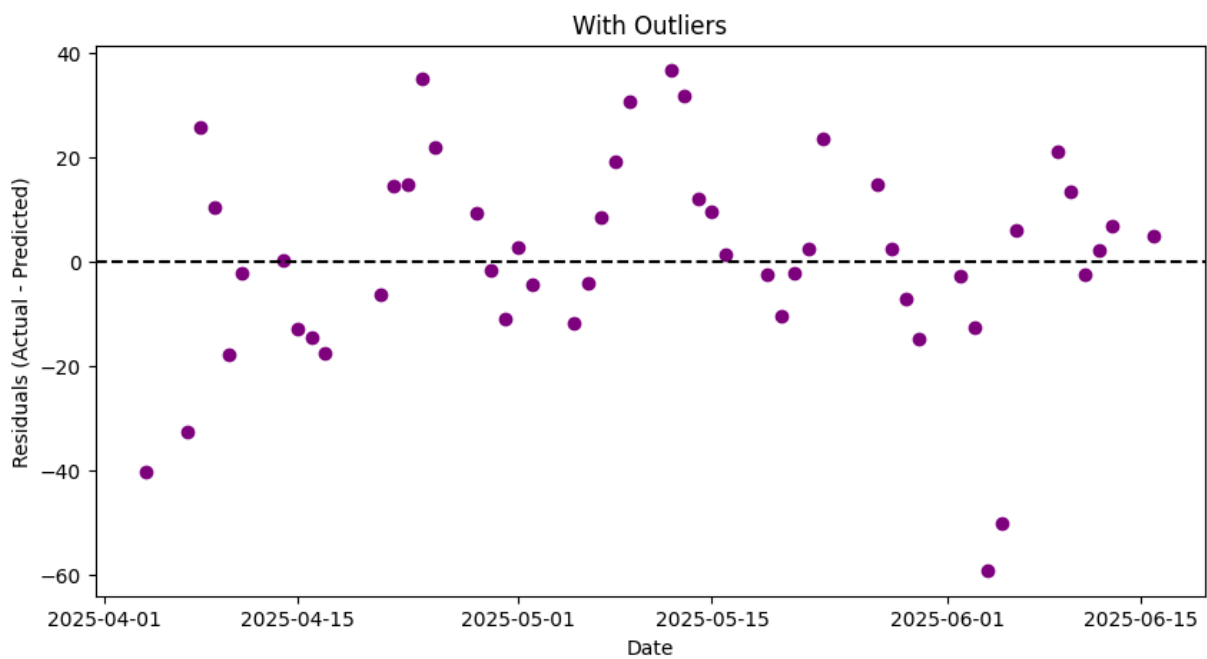
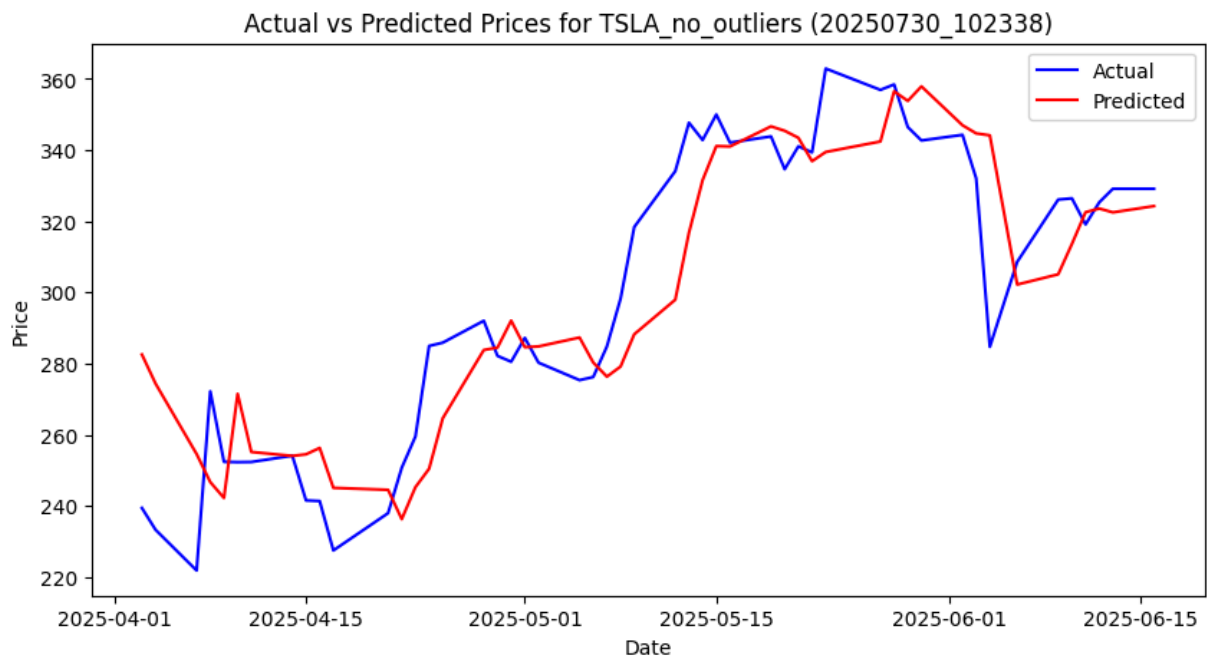
RMSE: 19.16

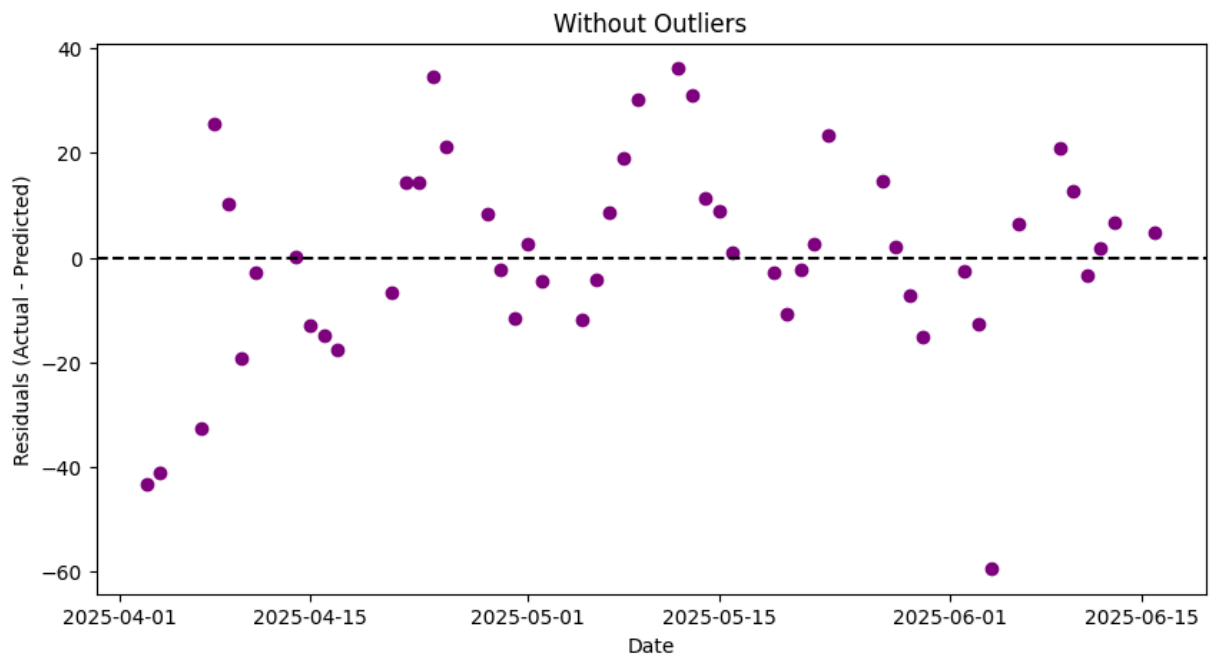
MAE: 14.30

R^2 : 0.78

Visualization







Summary of Findings

Model Performance

- **Variance Explained:** The linear regression models explain 77–78% of the variance in TSLA’s next-day closing price ($R^2 = 0.77$ with outliers, 0.78 without), demonstrating that features such as previous closing price (`prev_close`), trading volume (`volume`), and 5-day moving average (`ma5`) are effective predictors.
- **Prediction Accuracy:**
 - **Mean Absolute Error (MAE):** Approximately \$14 , indicating that typical predictions are reasonably accurate.
 - **Root Mean Square Error (RMSE):** Approximately \$19 , translating to a relative error of 5.24% to 8.56% across TSLA’s stock price range of \$221.86 to \$362.89 .
- **Relative Error Insights:** The model exhibits higher reliability at higher prices (e.g., 5.24% error at \$362.89) and lower precision at lower prices (e.g., 8.56% error at \$221.86), where the \$19 error is proportionally larger.

Model Comparison and Stability

- **Outlier Impact:** Predictions for June 5, 2025, showed minimal variation, with \$345.34 (with outliers) versus \$346.58 (without outliers)—a difference of

just \$1.24 . This small gap underscores the model's stability and low sensitivity to single outliers.

- **Residual Analysis:** Most residuals fall within $\pm \$40$, suggesting generally unbiased predictions. However, a significant overprediction of \$59.25 occurred on June 5, 2025 (predicted \$343.95 vs. actual \$284.70), due to a sharp 14.2% price drop (from \$332.05 to \$284.70), highlighting challenges in capturing abrupt market shifts.

Challenges and Limitations

- **Unexplained Variance:** Approximately 22–23% of price variation remains unaccounted for, likely influenced by external factors such as market sentiment or news events not captured by the current features.
- **Volatility Handling:** The model struggles with sudden volatility, as seen in the June 5, 2025, overprediction. This limitation stems from its reliance on historical lagged features, which may not signal rapid market changes effectively.

Practical Implications

- **Suitability:** With a relative error of 5.24–8.56% , the model may be suitable for long-term investment strategies, particularly at higher price levels. However, it is less reliable for short-term trading, especially during volatile periods or at lower stock prices, where the percentage error increases.

Recommendations and Next Steps

- **Model Enhancement:** In Phase 6, explore non-linear models (e.g., random forests) to better capture complex, non-linear patterns in the data.
- **Feature Expansion:** Incorporate volatility indicators in future iterations to improve the model's ability to predict during sudden market shifts.
- **Documentation:** Refine project documentation in Phase 8 to ensure scalability, clarity, and support for ongoing development.