

FinDraft: Analysis and Modeling of Anomaly Detection in Tesla Financial Data

Group 13: Ha LE, Mayra SUAREZ, Salam AL KAISSE

1 Business Problem

New retail investors often face difficulties understanding stock market trends and price fluctuations due to a lack of expertise. This knowledge gap hinders their confidence in starting investments, reduces their ability to make informed decisions, and increases their reliance on potentially unreliable sources.

The objective of this project is to bridge this gap by providing an automated system that generates clear and concise explanations of historical stock price trends. This solution will make the stock market more accessible to novice investors, empowering them with actionable insights.

2 Data Science Problem

To address the business problem, the following data science challenges are defined.

1. **Pattern Detection in Stock Price Data:** Analyze historical stock data to identify trends, patterns, and anomalies in price movements over various time frames (short-term, intermediate-term, long-term).
2. **Sentiment-Driven Contextualization:** Incorporate sentiment analysis of news articles to link significant price movements with corresponding news events, helping users understand the “why” behind price changes.
3. **Natural Language Generation (NLG):** Develop a system that translates complex stock market behaviors into easy-to-understand textual explanations, specifically tailored for non-technical users.
4. **Automation and Scalability:** Create a pipeline that dynamically retrieves stock data and relevant news, processes it, and automatically generates historical insights daily or over a customizable time range.

3 Data Collection and Preparation

- **Stock Data:** Historical stock data for Tesla and other top tech companies (2024) obtained from the Yahoo Finance API, including features such as open, high, low, close prices, and volume.
- **News Data:** News articles related to Tesla, including their title, description, publication date, sentiment scores, and sentiment labels, fetched from AlphaVantage API.

Both datasets can be extracted with a dynamic time range and updated daily.

4 Exploratory Data Analysis

4.1 Stock Data Insights

- The *average closing price* for Tesla during the analyzed period was \$248.42.
- Significant *price fluctuations* were observed during **October** and **November**, which coincided with key events such as *earnings reports* and *political developments*, reflecting how external factors can influence stock behavior.

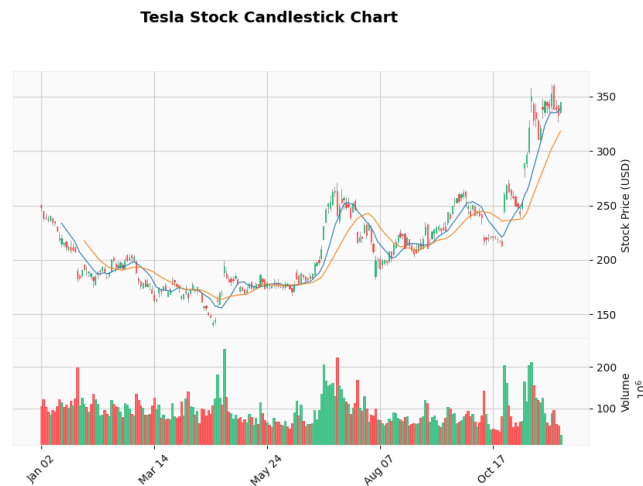


Figure 1: The candlestick chart reveals significant volatility during mid-June, coinciding with high investor activity and market speculation. Price fluctuations align with key news events highlighted later.

4.2 Sentiment Analysis

Positive sentiment scores correlated moderately with stock movements. The distribution of sentiment scores for Tesla-related articles is as follows:

- **Bearish:** Strong negative sentiment (below -0.3).
- **Somewhat-Bearish:** Mild negative sentiment (-0.3 to -0.1).
- **Neutral:** Balanced sentiment (-0.1 to +0.1).
- **Somewhat-Bullish:** Mild positive sentiment (+0.1 to +0.3).
- **Bullish:** Strong positive sentiment (above +0.3).

In total, 3,953 articles had positive sentiment, while 963 articles were negative, and only 22 articles were neutral.

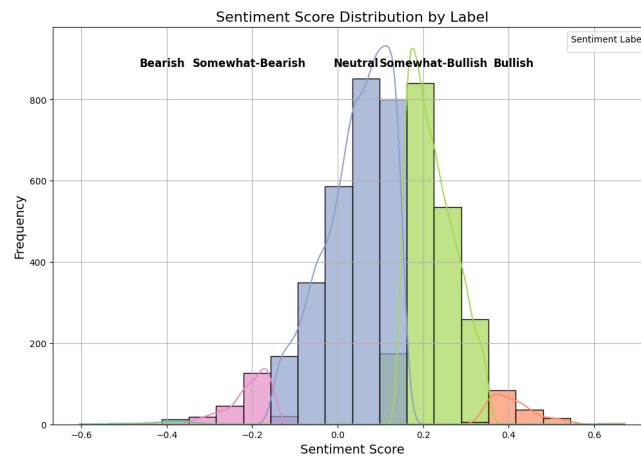


Figure 2: Most articles are Neutral, followed by Somewhat-Bullish, reflecting a generally balanced to optimistic outlook.

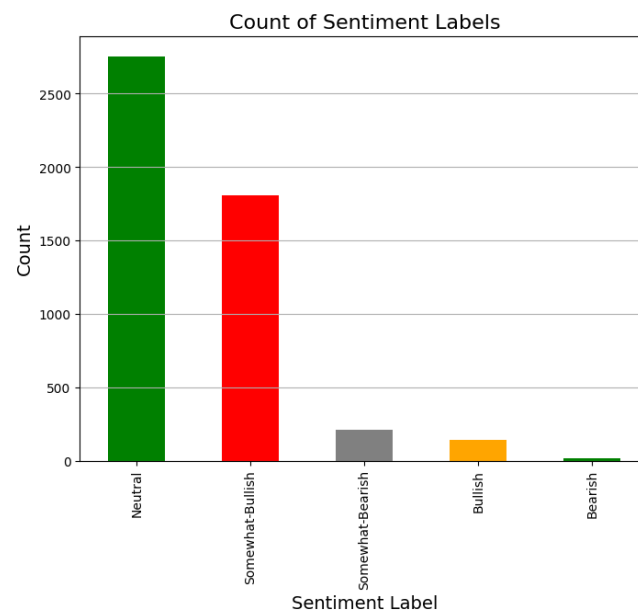


Figure 3: Count of sentiment labels.

5 Methodology

5.1 Trend Detection

Trends in stock prices are crucial for understanding market behavior and forecasting future movements. Trends can generally be classified based on their duration, as follows:

- **Short-term Trends:** Lasting hours to weeks, driven by daily fluctuations, news, or earnings reports.
- **Intermediate-term Trends:** Spanning weeks to months, influenced by economic factors like interest rates or industry developments.
- **Long-term Trends:** Extending months to years, shaped by fundamental factors such as economic growth or technological advancements.

The two primary types of trends are:

- **Uptrend:** Characterized by a series of higher highs and higher lows, indicating an overall upward movement in stock prices.
- **Downtrend:** Characterized by a series of lower highs and lower lows, signaling a decline in stock prices.

By applying the Savitzky-Golay filter to smooth the adjusted close price and using the `find_peaks` function from `scipy.signal`, we can detect peaks and troughs. Adjusting the conditions for continuous higher highs or lower lows, we can identify uptrends and downtrends for short- and intermediate-term periods, as shown in Figures 6 and 7.

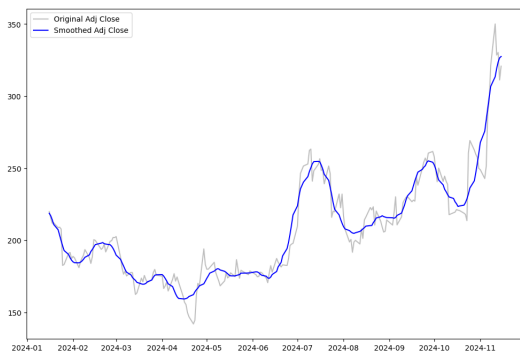


Figure 4: Smoothing the Adj Close using Savitzky-Golay filter

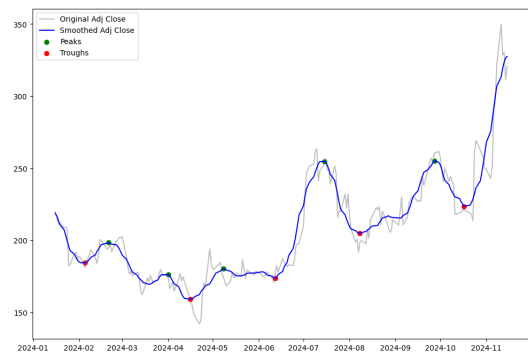


Figure 5: Detect peaks and troughs using `find_peaks()`

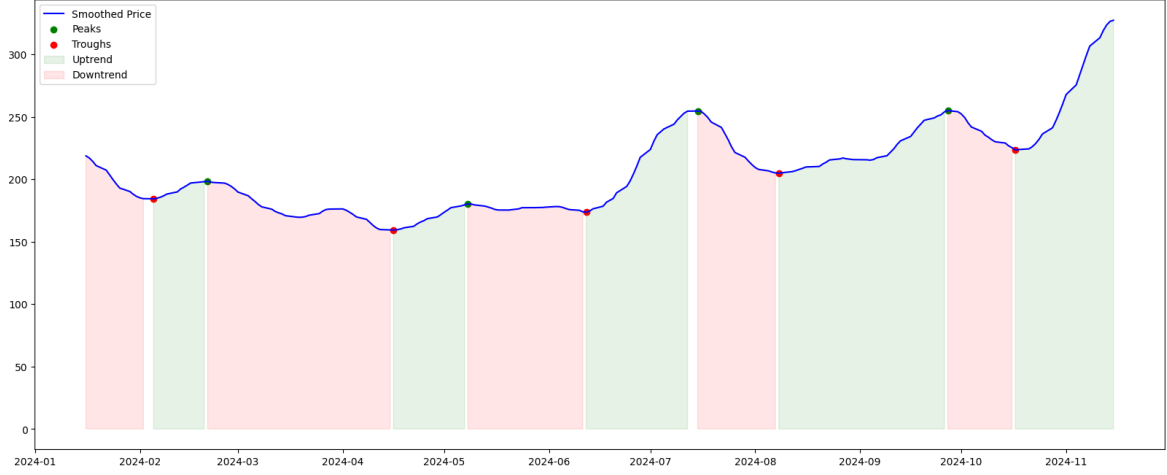


Figure 6: Short-term trends

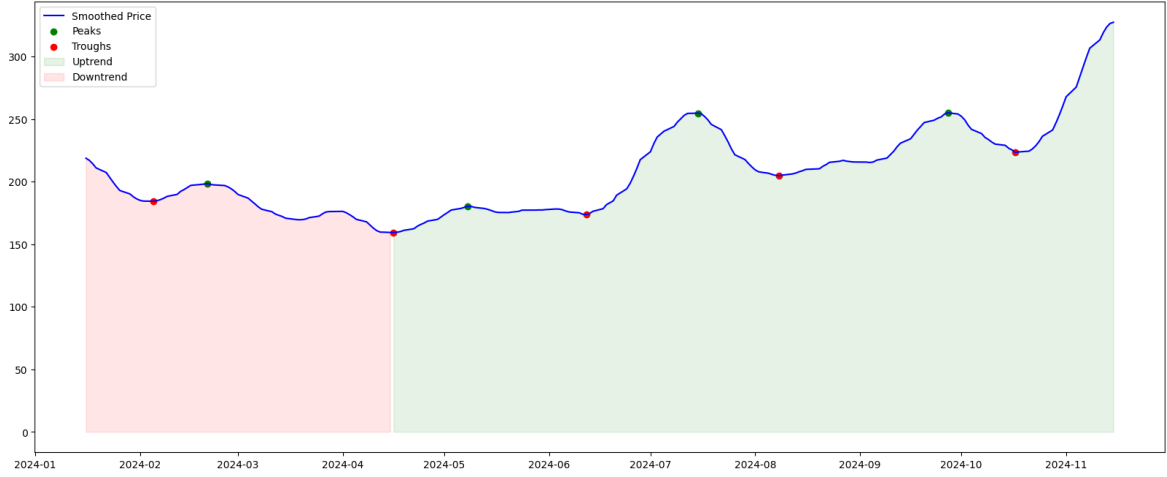


Figure 7: Intermediate-term trends

5.2 Anomaly Detection

5.2.1 Z-Score with Rolling Window:

The Z-score measures how many standard deviations a data point is from the mean. Using a rolling window approach, the Z-score is calculated over a moving subset of data, recalculating the mean and standard deviation for each window. This method captures short-term fluctuations and is ideal for detecting immediate market events and volatility.

$$Z_t = \frac{X_t - \mu_t}{\sigma_t}$$

Z-scores are typically used with two thresholds for detecting outliers:

- Z-Score > 2 or < -2 : Indicates moderate outliers, significantly different from the mean.

- **Z-Score** > 3 or < -3 : Indicates extreme outliers, far from the mean.

The Z-score is applied to stock price returns, which represent the percentage change in price over a period:

$$\text{Return} = \frac{P_t}{P_{t-1}}$$

where P_t and P_{t-1} are the current and previous prices.

Returns normalize price data, making it easier to compare assets and identify outliers, highlighting significant deviations in stock price behavior.

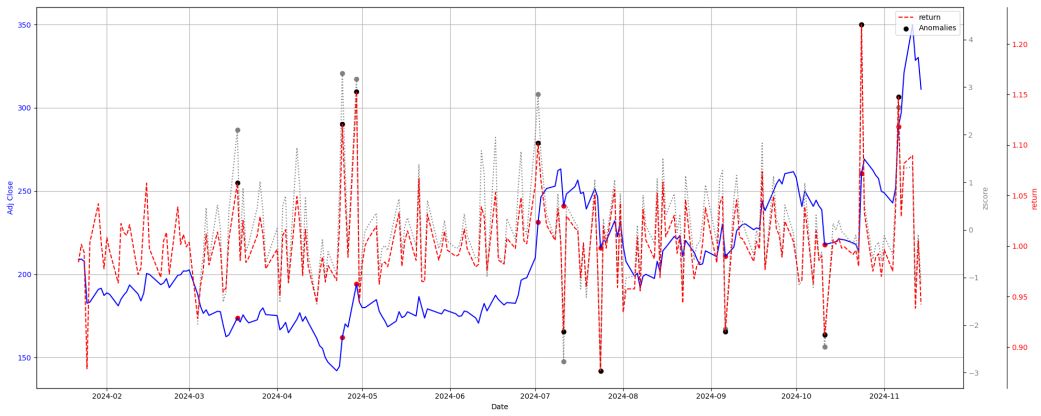


Figure 8: Anomalies detection using Z-score with `rolling_window = 30`, `threshold = 2`

Index	Date
1	2024-03-18
2	2024-04-24
3	2024-04-29
4	2024-07-02
5	2024-07-11
6	2024-07-24
7	2024-09-06
8	2024-10-11
9	2024-10-24
10	2024-11-06

Table 1: Dates of Detected Anomalies

5.2.2 Time Series Modeling

The ARIMA (AutoRegressive Integrated Moving Average) model is a statistical method used for time series forecasting and anomaly detection.

The graph of residuals reveals sharp spikes in November 2024, indicating unexpected movements in Tesla's stock price during this period. These spikes suggest that there were large deviations from the ARIMA(1, 1, 1) model's predictions, which could have been driven by external factors. The ACF and PACF plots show that the model did a good job capturing the short-term dependencies in the data. However, the presence of residual autocorrelation suggests that the model could still be improved by considering additional external variables or using a more complex model.

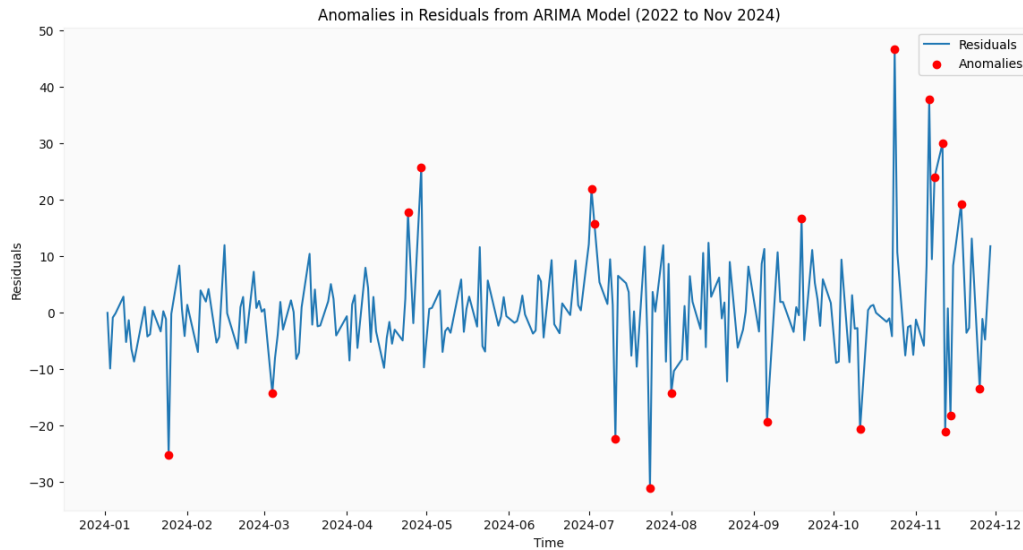


Figure 9: Graph of Tesla Adjusted Close Price and Sentiment Deviation with Anomalies in ARIMA Model.

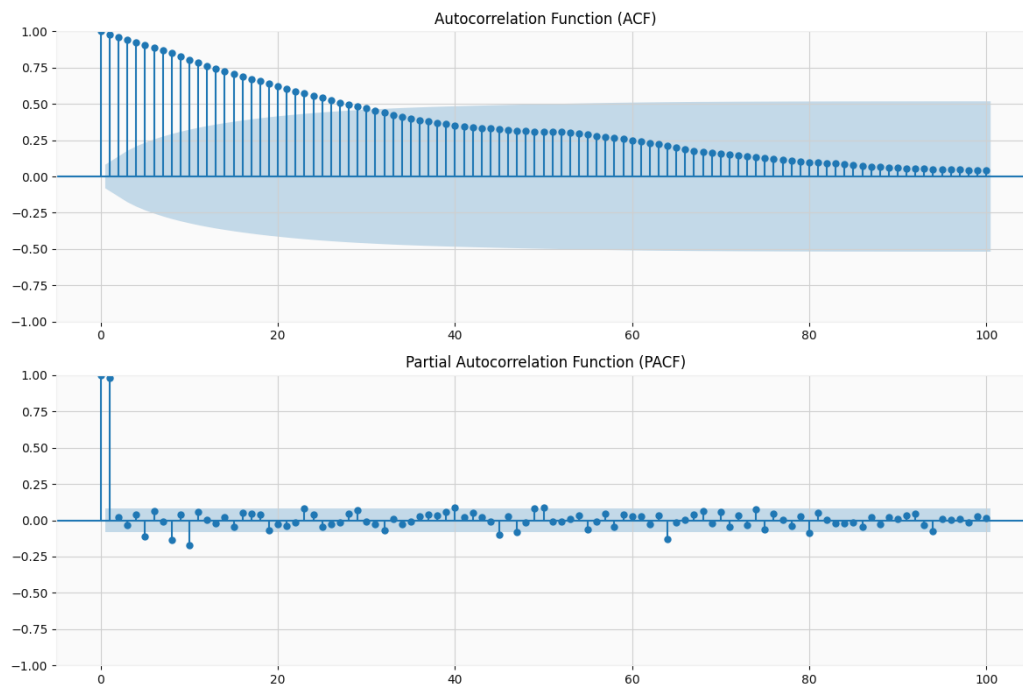


Figure 10: Graph of ACF and PACF.

5.3 Sentiment data and Keyword Analysis

- Extracted and categorized keywords from news headlines. Keywords such as "tesla", "tsla" and "Elon"

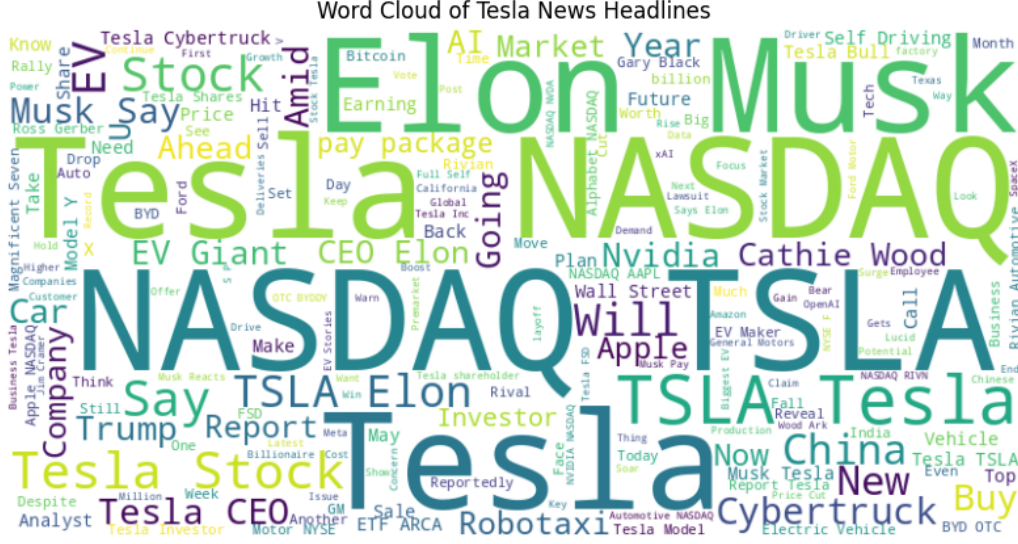


Figure 11: Word Cloud of Tesla News Headlines.

- Performed sentiment data correlation analysis with Tesla stock price.

To analyze the correlation between the sentiment score and the adjusted close price, we decided to use the 7-day sentiment deviation, which is the difference between the 7-day rolling average sentiment and the overall average sentiment (baseline).

$$\text{7-day sentiment} = \frac{1}{7} \sum_{i=t-6}^t \text{sentiment_score}_i$$

$$\text{baseline sentiment} = \frac{1}{n} \sum_{i=1}^n \text{sentiment_score}_i$$

$$\text{sentiment deviation} = \text{7-day sentiment} - \text{baseline sentiment}$$

The reasons for using 7-Day sentiment deviation are :

- **Noise reduction:** A 7-day window smooths out daily fluctuations, focusing on longer-term sentiment trends.
- **Trend detection:** It captures sustained sentiment shifts, better reflecting the impact of sentiment on stock prices over time.
- **Outlier minimization:** It reduces the effect of short-term sentiment spikes, ensuring that only meaningful sentiment changes influence the analysis.
- **Market reaction alignment:** Stock prices typically respond to sentiment over days or weeks, making the 7-day deviation more relevant for capturing price movement influences.

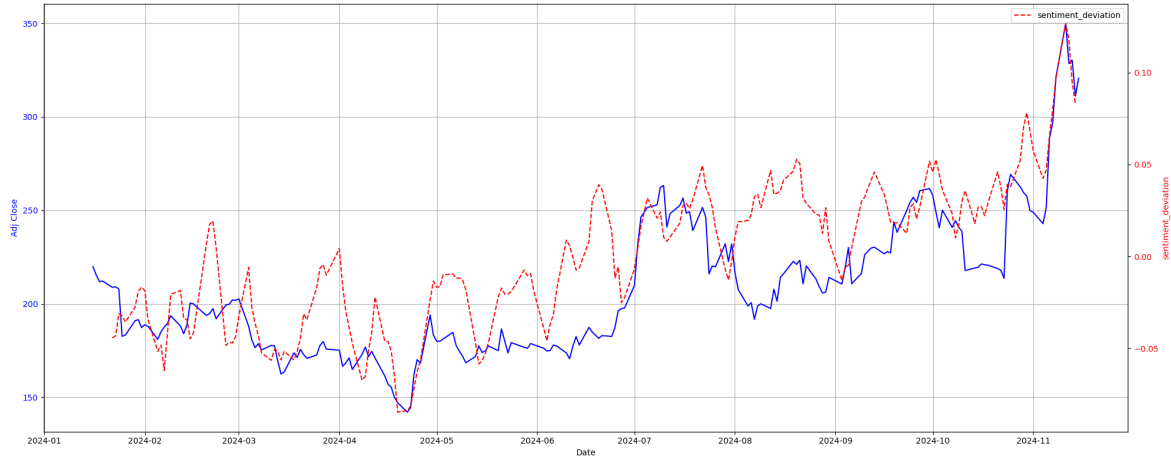


Figure 12: Intermediate-term trends

The Pearson correlation coefficient between Tesla's adjusted stock price and sentiment deviations is 0.819, with a p-value of 1.13×10^{-51} . This suggests a strong positive linear relationship between the two variables, with the p-value indicating that the correlation is highly statistically significant.

6 Results

Z-Score-based anomaly detection and ARIMA were used to identify significant outliers in Tesla's stock price during the observed period, enabled us to pinpoint unusual spikes and dips in the stock.

Additionally, sentiment analysis was employed to assess public perception of Tesla during these events. By analyzing news articles and social media posts, we tracked sentiment trends and matched them with stock movements. The results clearly indicated that positive news sentiment was closely linked to stock price increases.

These methods flagged anomalies on October 23rd and November 12th, aligning with major news events:

- October 23rd: Driven by negative sentiment aligning with news events reporting concerns about Cybertruck's recall.
- November 12th: Driven by positive sentiment, this spike aligns with the re-election of Donald Trump as U.S. President and his appointment of Elon Musk to lead the Department of Government Efficiency (DOGE).

Figure 13 highlights the anomalies detected using Z-Score with a lower threshold to minimize the risk of missing significant abnormal events. It also illustrates a linear relationship between sentiment deviation and the adjusted close price.

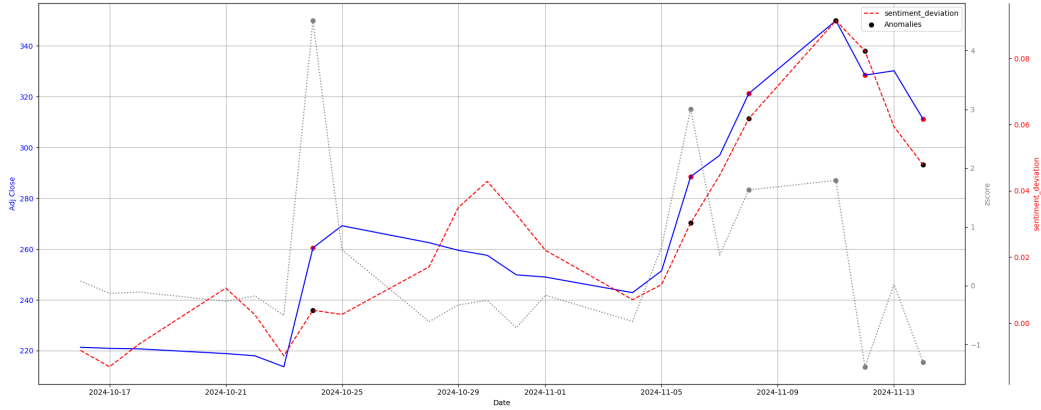


Figure 13: Adj Close and Sentiment with anomalies

Below are other key findings from the analysis:

- Anomalies in stock prices were closely linked to significant Tesla-related news events, such as Elon Musk’s political role and product recalls.
- Positive sentiment keywords like ”EV”, ”robotaxi”, and ”Musk” often preceded stock price surges.
- Sentiment analysis revealed that the majority of articles were classified as Neutral, followed by Somewhat-Bullish and Bearish, reflecting a predominantly balanced to optimistic outlook.

7 Generated Report Outputs

7.1 Actionable Insights

Based on the findings from the sentiment analysis and anomaly detection, we have identified several strategies that investors can apply to better understanding of market signals :

- Investors could monitor sentiment trends and high-impact keywords in news articles to anticipate upward stock movements.
- Detecting price anomalies around major news events provides strategic entry and exit points for market participants.
- Implementing real-time sentiment tracking could allow investors to respond promptly to market-moving news and shifts in sentiment
- Articles published on the day of anomaly events, as well as three days before and after, should be analyzed to assess their impact on stock prices.
- Automating the report generation process would improve the efficiency of anomaly detection and sentiment tracking.

<https://www.benzinga.com/government/regulations/24/11/41867551/tesla-faces-nhtsa-scrutiny-over-social-media-promotion-of-fsd-technology-that-shows-disen>

- Tesla Remains Core For Gary Black Despite Price Cuts, Earnings Drop - Nvidia, Google, Eli Lilly Among Top Performers in Tech Bull's Portfolio - The Future Fund Active ETF (ARCA:FFND) - Salesforce (NYSE:CRM) (2024-11-11): Gary Black, the managing partner of The Future Fund, still considers Tesla Inc. (TSLA) a key investment despite concerns about the company's recent price cuts for its electric vehicles and their potential effect on earnings. Tesla is a significant holding in Black's portfolio, although he has reduced the position due to these worries. In addition to Tesla, Black's top performers include Nvidia, Google, and Eli Lilly in the tech sector. The Future Fund Active ETF (ARCA:FFND) and Salesforce (Read more: <https://www.benzinga.com/analyst-ratings/analyst-color/24/11/41870590/tesla-remains-core-for-gary-black-despite-price-cuts-earnings-drop-nvidia-google-el>)

- Tesla Shorts Suffer \$5 Billion Blow After Trump's Election: Report - Tesla (NASDAQ:TSLA) (2024-11-11): The article discusses the financial impact of Donald Trump's election on hedge funds that had short positions in Tesla. After Trump's victory, these hedge funds incurred substantial losses totaling \$5.2 billion. This event highlights the risks associated with short-selling, where investors bet against a company's stock, hoping to profit from a decline in its price. However, in this case, the election outcome seems to have influenced the market sentiment, resulting in a negative impact on the short-sellers' positions. (Read more: <https://www.benzinga.com/markets/equities/24/11/41870874/tesla-shorts-suffer-5-billion-blow-after-trumps-election-report>)

- Cybertruck 'Likely Won't Be Profitable' In Q4, Says Tesla Researcher - Tesla (NASDAQ:TSLA) (2024-11-11): The article suggests that Tesla's Cybertruck is unlikely to be profitable in the fourth quarter, according to researcher Troy Teslike. This is because Tesla has started selling the regular versions of the Cybertruck instead of the Foundation series, which may impact profitability. (Read more: <https://www.benzinga.com/tech/24/11/41873767/cybertruck-likely-wont-be-profitable-in-q4-says-tesla-researcher>)

- Tesla Bull Hikes Elon Musk's EV Giant's Price Target To \$400 As Stock Rallies 47% In One Month - Tesla (NASDAQ:TSLA) (2024-11-11): The article discusses Wedbush analyst Dan Ives' decision to raise Tesla's price target from \$300 to \$400, citing the potential impact of Donald Trump's election win on the company's success. This optimism comes as Tesla's stock has already experienced a significant 47% rally in just one month. The article highlights the positive sentiment surrounding Elon Musk's EV company and the analyst's belief that Trump's victory could be a significant turning point for Tesla's future prospects. (Read more: <https://www.benzinga.com/analyst-ratings/price-target/24/11/41875624/tesla-bull-hikes-elon-musk-s-ev-giants-price-target-to-400-as-stock-rallies-47-in-one>)

- Tesla Stock Gains Over 7% In Monday Pre Market: What's Going On? - Tesla (NASDAQ:TSLA) (2024-11-11): The article discusses the recent surge in Tesla's stock price, which gained over 7% in pre-market trading on Monday. This increase follows Tesla's achievement of a \$1 trillion market capitalization on Friday, a significant milestone for the electric vehicle company. The stock's upward trend is attributed to Tesla's strong performance and its position as a leading electric vehicle manufacturer. (Read more: <https://www.benzinga.com/tech/24/11/41875643/tesla-stock-gains-over-7-in-monday-pre-market-whats-going-on>)

- Tesla, Inc. (TSLA) is Attracting Investor Attention: Here is What You Should Know (2024-11-11): Tesla, Inc. (TSLA) has been attracting significant attention from investors, as indicated by the increased interest from Zacks.com users. Here are some key points to consider regarding Tesla:

1. Electric Vehicle (EV) Leadership: Tesla is a pioneer and leader in the electric vehicle market, known for its innovative technology and sleek designs. Its focus on sustainable transportation has positioned it as a key player in the global shift towards electric mobility.

2. Product Lineup: Tesla offers a (Read more: <https://www.zacks.com/stock/news/2368004/tesla-inc-tsla-is-attracting-investor-attention-here-is-what-you-should-know>)

- Archer Aviation Stock and Joby Aviation Stock: Are They Ready for Takeoff? (2024-11-11): Archer Aviation and Joby Aviation, two electric aircraft manufacturers, are poised for potential growth and success in the market, similar to the early days of Tesla. Both companies have recently gone public and are developing electric vertical takeoff and landing (eVTOL) aircraft for urban air mobility. Archer Aviation, backed by United Airlines, has a market capitalization of around \$1.2 billion and aims to begin commercial operations in 2025. Joby Aviation, backed by Toyota and Uber, has (Read more: <https://www.fool.com/investing/2024/11/11/archer-aviation-stock-and-joby-aviation-stock/>)

- Tesla shares pop 8% as postelection rally continues (2024-11-11): The article highlights that Tesla's stock price experienced a significant increase of approximately 8% on Monday, following the election of President-elect Trump. This rise in share price is attributed to the ongoing post-election rally as investors respond positively to the initial developments in Trump's presidential cabinet formation. (Read more: <https://www.cnbc.com/2024/11/11/tesla-shares-pop-8percent-as-postelection-rally-continues-donald-trump-elon-musk.html>)

- Should You Buy Tesla ETFs on \$1 Trillion Comeback? (2024-11-11): The article discusses the recent surge in Tesla's market value, surpassing \$1 trillion again, and explores whether investing in Tesla ETFs is a good decision. The primary reasons for Tesla's comeback are attributed to Donald Trump's presidential election victory and the company's impressive Q3 2024 results. The article suggests that Trump's win has positively influenced the stock

Figure 16: Generated Report Page 3

8 End User Testing and Hypotheses

8.1 Hypotheses

In this analysis, the following hypotheses were proposed and tested:

- Positive sentiment in news data correlates moderately with upward stock price movements, particularly during Tesla's peak periods, such as post-election surges and product announcements.
- Anomalies detected in stock price data often coincide with significant market or company-specific events, such as earnings reports, political events, or major product announcements. These anomalies provide critical signals for investors.
- Sentiment-driven keyword analysis is a reliable indicator for analyzing stock trends compared to generic news sentiment.

8.2 Discussion

Limitations include the assumption of sentiment analysis accuracy and potential dataset biases. Thresholds used in anomaly detection need further review and validation by financial experts to ensure robustness.

8.3 End User Testing

To validate the practical applicability of our analysis, the generated report was presented to retail investors for usability and interpretability testing. Key findings from the testing include:

- Users found the combination of historical stock market data and sentiment analysis insightful for identifying market trends and anomalies.
- Feedback emphasized the clarity of visualizations, such as candlestick charts and word clouds, which made the analysis more digestible and actionable.
- Suggestions included incorporating real-time updates and expanding datasets to include additional companies or industries for broader applicability.

9 Conclusion and Future Work

This analysis successfully demonstrated the power of combining historical stock price trends with sentiment analysis to detect anomalies and identify key market movements. The ability to identify these anomalies provides valuable insights into the dynamics of stock behavior and investor sentiment. However, there are several areas for future work:

- Expanding datasets to include stock data from other tech companies or industries would broaden the scope and applicability of this approach.
- Improving sentiment models through machine learning could refine the accuracy of sentiment predictions and anomaly detection, providing more reliable real-time insights.

Overall, the integration of sentiment analysis with stock price data has proven to be a promising tool for detecting market anomalies and providing investors with actionable insights. The next steps will focus on scaling the model, enhancing its accuracy, and applying it to real-world trading scenarios.

References

- Reichenbach, Felix. Exploring the Intersection of Behavioral and Digital Finance: Essays on Crowdfunding, Social Media, Stock Charts, and the Behavior of Retail Investors. 2024, Technische Universität Berlin This dissertation delves into behavioral finance, examining how sentiment from platforms like social media and stock forums interacts with financial trends. It provides theoretical grounding for the role of sentiment in understanding historical financial behavior,
- Mamun, Abdullah Al, et al. "Machine Learning for Stock Market Security Measurement: A Comparative Analysis of Supervised, Unsupervised, and Deep Learning Models." International Journal of Networks and Security, vol. 4, no. 1, 2024, pp. 22–32, <https://doi.org/10.55640/ijns-04-01-06>.

this paper focuses on comparing machine learning techniques for stock market analysis, its use of sentiment data as a critical input aligns with our project's emphasis on the relationship between historical sentiment trends and financial data