Predicting Post-Election FAANG Stock Prices: Accurate November Trends and Flat December Projections*

An XGBoost Model Analysis of Market Behavior and Real-World Implications for Meta, Amazon, Apple, Netflix, and Google

Jimin (Jamie) Lee

December 3, 2024

This paper presents a predictive model for FAANG stock prices (Meta, Amazon, Apple, Netflix, and Google) during November and December 2024. Using historical stock price data from Yahoo Finance and an XGBoost regression framework, we estimate adjusted stock prices based on features such as lagged prices, moving averages, and volatility. Our findings indicate strong predictive accuracy for short-term trends in November, with predictions closely aligning with actual prices for stocks like AAPL and GOOGL. However, discrepancies, such as the underestimation of NFLX prices, highlight the impact of unique stock-specific dynamics. Predictions for December show a flat trajectory, potentially influenced by data limitations and post-election market sentiment. This analysis provides valuable insights into FAANG stock behavior during critical periods, equipping investors and policymakers to better understand market responses to external events like elections.

Table of contents

---

*Code and data are available at: https://github.com/jamiejiminlee/FAANG-Stock-Forecast.git.

1 Introduction

The performance of FAANG stocks - **Meta (META)**, **Amazon (AMZN)**, **Apple (AAPL)**, **Netflix (NFLX)**, and **Google (GOOGL)**, has become a cornerstone of global financial markets, reflecting broader economic trends and investor sentiment. These companies represent diverse sectors, from e-commerce and technology to entertainment and digital advertising, making their stock price movements critical indicators of market health. In the context of the 2024 post-pandemic recovery and a pivotal United States Presidential Election, understanding and predicting FAANG stock price trends is particularly relevant for investors and policymakers alike.

This paper focuses on historical stock price data from years 2020 to 2024, sourced from Yahoo Finance (Team 2024) to develop predictive models for FAANG stock prices in November and December 2024. Utilizing XGBoost regression models, we analyze historical trends and engineered features, including lagged prices, moving averages, and volatility measures, to forecast stock price movements. These predictions provide valuable insights into the behavior of FAANG stocks during critical periods, such as post-election market adjustments, and highlight opportunities and risks for short-term and long-term investment strategies.

The primary estimand of this study is the predicted adjusted stock price for each FAANG company during November and December 2024, derived from features constructed using his-

torical price trends and technical indicators. Our findings reveal that predictions for November align closely with observed trends for most stocks, reflecting the model's ability to capture short-term dynamics. However, discrepancies, such as the underestimation of NFLX prices, underscore the influence of unique stock-specific factors. For December, predictions indicate a flat trajectory across stocks, likely influenced by data limitations and post-election market sentiment.

The implications of these findings extend beyond price predictions, offering a framework to evaluate market behavior in response to external factors such as elections and economic conditions. The remainder of this paper is organized as follows: Section 2 describes the data sources, variables, and preprocessing steps; Section 3 outlines the modeling approach and feature engineering strategies; Section 4 presents the results and discusses trends in predicted stock prices; and Section 5 addresses the implications and limitations of the findings, with additional model diagnostics detailed in the Appendix B.

2 Data

2.1 Data Overview

This analysis uses historical stock price data for FAANG companies (Meta, Amazon, Apple, Netflix, and Google) obtained from the Yahoo Finance API (Team 2024) via the `tidyquant` of Dancho (2024) library in R (R Core Team 2023) on November 30, 2024. The data ranges from January 1, 2020, to November 29th, 2024, providing daily records of stock performance, including key metrics such as opening, high, low, and closing prices, trading volume, and adjusted closing prices normalized for corporate actions. To extend the data for December 2024, placeholder dates are generated and combined with the existing data, followed by interpolation to fill missing values in the adjusted price column. Such data cleaning ensures a complete and continuous dataset for model training and prediction. Details of the data cleaning and interpolation process are provided in Appendix A.1.

Data cleaning and analysis are conducted by employing the following packages: 'tidyverse' of Wickham and contributors (2024), 'TTR' of Ulrich (2024), and 'arrow' of Foundation (2024).

2.2 Measurement

Stock prices reflect a company's perceived value in financial markets, influenced by factors like economic conditions, corporate performance, and investor sentiment. These dynamics are captured through daily trading data, where variables such as **'open'**, **'high'**, **'low'**, **'close'**, **'adjusted'**, and **'volume'** provide a structured snapshot of market activity. Each data point represents an aggregated view of trading outcomes for a given day, simplifying continuous price fluctuations into discrete records. For example, the adjusted price accounts for corporate

actions like stock splits and dividends, offering a standardized metric for historical comparison. Volume indicates total shares traded, reflecting market liquidity and investor interest but not the motivations behind trades.

The data assumes consistency across trading days, where daily aggregates like close encapsulate the final market consensus for that day. However, this measurement approach may overlook intraday volatility and the nuanced impact of macroeconomic events or breaking news that influence trading behavior. Similarly, variables like date simplify time into discrete intervals, capturing trends over time but not the continuous nature of market dynamics. While these measurements are essential for analyzing historical trends, they simplify complex phenomena like investor sentiment and market reactions, such as the rationale behind trades or varying confidence levels. To address these limitations, feature engineering constructs variables that capture trends, volatility, and other patterns, enhancing the data's analytical value for model training and predictive analysis. Further details are provided in Appendix A.

2.3 Variables

The collected data from Yahoo Finance (Team 2024) contains several key variables relevant to the analysis.

- **symbol**: The stock ticker symbol identifying the company (e.g., AAPL for Apple, AMZN for Amazon).
- **date**: The trading date, essential for analyzing time-series trends.
- **open**: The stock's opening price for the trading day, indicating initial market conditions.
- **high**: The highest price reached during the trading day, reflecting intraday volatility.
- **low**: The lowest price during the trading day, reflecting downward market trends.
- **close**: The stock's unadjusted closing price, representing the final value at the end of trading.
- **volume**: The total number of shares traded during the day, reflecting market activity.
- **adjusted**: The closing price adjusted for corporate actions such as splits or dividends, ensuring consistency for trend analysis.

2.3.1 Constructed Variables

To enhance the data's analytical depth and support predictive modeling, the following constructed variables were engineered to capture price trends, volatility, and momentum:

- **Lag_1**: The closing price from the previous trading day, derived using the `lag()` function. This variable captures sequential patterns in stock prices.
- **Rolling_Mean_7**: A 7-day moving average of adjusted closing prices, computed using the `rollmean()` function. This variable smooths short-term fluctuations to highlight weekly trends.

- **sma_20**: The 20-day simple moving average (SMA) of adjusted prices, reflecting medium-term price trends.
- **sma_50**: The 50-day simple moving average of adjusted prices, providing insights into longer-term trends.
- **volatility**: The 20-day rolling standard deviation of adjusted prices, capturing short-term price variability.
- **daily_return**: The percentage change in adjusted closing prices relative to the previous day. This value is calculated by dividing the adjusted closing price by its value from the previous day and subtracting 1. This variable normalizes price changes, making them comparable across time and stocks.
- **symbol_encoded**: A numeric encoding of the stock ticker symbol, allowing the categorical **symbol** variable to be included in the modeling process.

2.3.2 Outcome Variable

The **outcome variable** for this analysis is the **adjusted closing price (adjusted)**, which serves as the target for prediction. This represents the stock's closing price adjusted for corporate actions like stock splits or dividends, providing a normalized measure of stock value. The model aims to predict this variable based on historical data and constructed predictors.
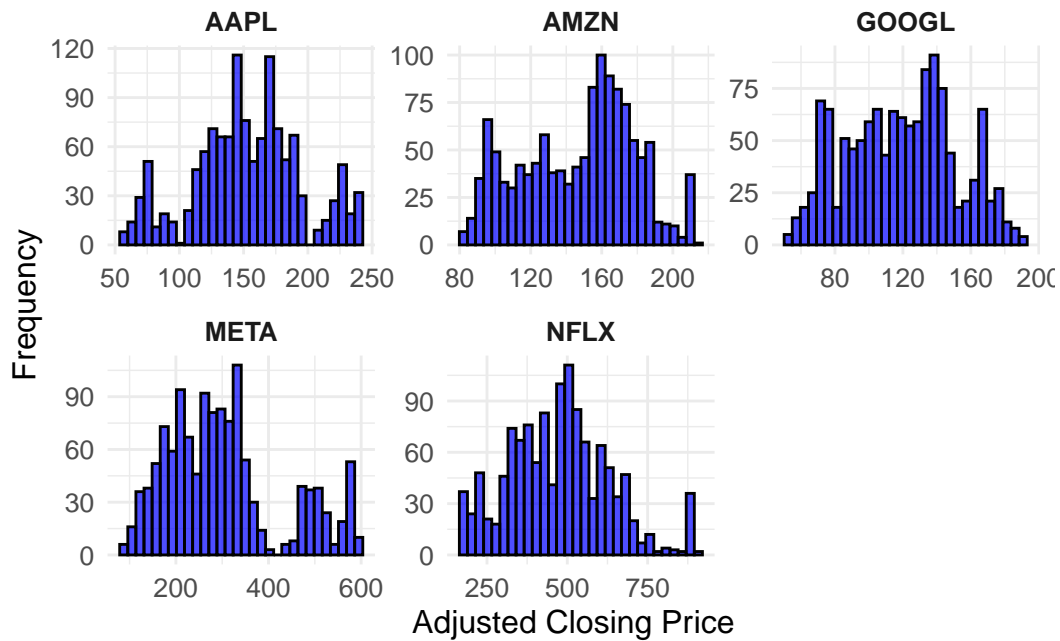

Figure 1: Distribution of Adjusted Closing Prices by Stock

Figure 1 illustrates the distribution of the adjusted closing prices (**'adjusted'**) for each FAANG

stock, faceted by stock symbol. These distributions highlight the variability in stock prices across the companies, with stocks like NFLX showing higher price ranges, whereas GOOGL and AMZN have narrower distributions. The visual emphasizes the diverse price behavior among FAANG stocks, underscoring the need for tailored models that capture these unique trends. By normalizing for corporate actions like stock splits or dividends, the adjusted closing price serves as a reliable target variable for forecasting and analysis.

2.3.3 Predictor Variables

The predictor variables used to model the adjusted closing price are:

- `Lag_1`: The prior day's adjusted closing price, reflecting momentum.
- `Rolling_Mean_7`: A smoothed metric capturing weekly price trends.
- `sma_20` and `sma_50`: Simple moving averages representing medium- and long-term price trends.
- `volatility`: The short-term price variability, highlighting potential risks.
- `daily_return`: The normalized daily percentage price change.
- `symbol_encoded`: A numerical representation of the stock ticker symbol, capturing company-level differences.

Together, these variables provide a robust framework for understanding and predicting stock price movements over time, utilizing both raw market indicators and derived metrics.

Table 1: Random Sample of Predictive Variables for Modeling Adjusted Closing Price

Lag_1	Rolling Mean (7-Day)	SMA (20-Day)	SMA (50-Day)	Volatility	Daily Return	Symbol (Encoded)
186.33	190.14	185.31	179.13	6.63	-0.01	2
207.89	207.89	205.95	196.42	4.18	0.00	2
125.17	128.87	128.21	133.18	2.43	0.01	2
171.96	163.05	166.76	167.60	6.14	0.00	1
305.92	319.23	331.31	330.76	12.26	0.04	4

Table 1 displays a random selection of five rows from the dataset, showcasing the predictive variables utilized in the modeling process. Each row represents data for a specific stock, including values for lagged adjusted closing prices (Lag_1), smoothed weekly averages (Rolling Mean (7-Day)), and medium- and long-term moving averages (SMA (20-Day) and SMA (50-Day)). These variables are complemented by measures of price volatility, daily percentage returns (Daily Return), and an encoded numerical identifier for the stock symbol (Symbol (Encoded)). All values are rounded to two decimal places to enhance clarity and precision.

3 Model

3.1 Model Overview

The XGBoost regression models in this study predict adjusted closing prices for FAANG stocks—Meta, Amazon, Apple, Netflix, and Google—for November and December 2024. Separate models are constructed for each month to account for time-specific market dynamics and recent trends. To capture short and long-term trends in stock price movements, these models incorporate features such as lagged prices, moving averages, daily returns, and volatility. Training data is segmented by timeframe, with the November model using data up to October 31, 2024, and the December model using data through November 30, 2024. Temporal features and technical indicators enable the models to capture stock-specific volatility and broader market trends. Further details on model specifications and diagnostics are provided in Appendix B.

3.2 Model Assumptions

To ensure the validity and reliability of our XGBoost models, several key assumptions are made during the modeling process:

- **Non-Linearity in Price Movements**: Stock prices often exhibit non-linear patterns due to various market forces, including economic news, company performance, and broader market trends. To account for these non-linearities, we use XGBoost, which allows for the modeling of complex, non-linear relationships between the predictors (such as lagged prices and moving averages) and the target variable (adjusted closing prices). This assumption ensures that the model can effectively capture the changing dynamics of stock price movements over time.

- **Stationarity of Price Features**: The model assumes that the selected technical indicators—such as moving averages, volatility, and daily returns—are stationary over the forecasting period. While stock prices are inherently volatile, we assume that the patterns observed in historical data are stable enough to provide reliable predictions for the future months. This assumption allows us to model stock prices without explicitly accounting for trends or seasonality, assuming that the past behavior is a good representation of future dynamics.

- **Independence of Observations**: Each daily stock price observation is assumed to be independent of the others. While stock prices are influenced by previous values (as captured through lagged features), we assume that each day's stock price data point does not directly depend on other data points in ways not captured by the lagged variables or technical indicators. This assumption is important for simplifying the modeling process and ensuring that the features can be treated as independent predictors.

- **Stationarity of Model Parameters**: We assume that the relationships between the stock prices and the chosen technical indicators (e.g., moving averages, volatility) remain relatively stable over the forecast horizon. While stock prices can be influenced by many unpredictable factors, this assumption allows the model to generalize well to the forecasting periods (November and December 2024).

- **Feature Importance**: We assume that the selected features—lagged prices, moving averages, volatility, and daily returns—capture the most important aspects of stock price movements for the FAANG companies. While other factors (such as macroeconomic events, news sentiment, or geopolitical developments) may influence stock prices, these features are chosen for their relevance to technical analysis and their ability to capture short-term and long-term market trends.

- **Use of XGBoost's Flexibility**: The XGBoost algorithm is assumed to be an appropriate method for this analysis due to its ability to handle large datasets, capture non-linear relationships, and deal with complex interactions between features. We assume that its regularization techniques will help prevent overfitting while still capturing the intricate patterns present in the stock price data.

3.3 Model Setup

To accommodate differences in data availability and prediction requirements, the modeling process was split into two separate models for November and December. The November model was trained and validated using historical data up to October 31, 2024, enabling predictions that could be compared to actual observed values for November. In contrast, the December model was trained on data up to November 30, 2024, and relied solely on historical trends and engineered features for prediction, as actual December data was unavailable. This separation allowed the analysis to leverage observed data for evaluating short-term accuracy in November while extending the prediction framework to December despite the lack of real data.

Modeling processes are conducted by employing the following packages: `tidyquant` package of Dancho (2024) for downloading and managing financial data, `xgboost` package of Chen and Guestrin (2024) for implementing the machine learning model, `dplyr` package of Wickham et al. (2024) for data manipulation, and `lubridate` package of Grolemund and Wickham (2024) for handling dates and times.

3.3.1 November Model

The primary estimand of the November model is the adjusted closing price of FAANG stocks for November 2024. The model is structured as a regression task using the XGBoost algorithm to estimate these prices for each of the five FAANG companies. The model setup is as follows:

$$\hat{y}_i = f(X_i)$$

Where:

- $\hat{y}_i$ is the predicted adjusted closing price for the stock on day i.
- X_i represents the feature set for day i, which includes:

  - `Lag_1`: The adjusted closing price of the previous day.
  - `Rolling_Mean_7`: The 7-day rolling average of the adjusted closing price.
  - `sma_20`: The 20-day simple moving average.
  - `sma_50`: The 50-day simple moving average.
  - `volatility`: The 14-day rolling standard deviation of daily returns.
  - `daily_return`: The daily percentage change in the adjusted closing price.
  - `symbol_encoded`: A numeric encoding for the FAANG companies (Meta, Amazon, Apple, Netflix, Google).

The data for the November model is prepared by extracting the relevant data from January 2020 to October 31, 2024. Features like moving averages and volatility are calculated based on this historical data, and the model is trained to predict the adjusted closing prices for November 2024.

The model was implemented using the `xgboost` of Chen and Guestrin (2024) in R, which is optimized for performance in regression tasks:

$$y_i = \text{XGBoost}(X_i, \theta)$$

Where y_i represents the actual adjusted closing price for day i and θ represents the model parameters to be learned during training.

Key hyperparameters for the model are specified as:

- **Max Depth**: 6, to control the complexity of the trees and prevent overfitting
- **Learning Rate**: 0.1, to balance the contribution of each boosting iteration
- **Number of Rounds**: 100, to determine the total iterations for model optimization
- **Objective Function**: Regression with squared error, suitable for continuous target variables.

This set-up ensures that the model captures patterns in stock price movements while balancing bias and variance for improved predictive performance. Key assumptions include the appropriateness of historical stock price trends and technical indicators for predicting future prices.

3.3.2 December Model

The December model follows a similar approach to the November model but incorporates extrapolated data for December 2024. Placeholder dates for December are added, and missing values in the adjusted price column are filled using forward and backward interpolation. This ensures the model reflects recent trends while accounting for the data cutoff on November 30, 2024. The model structure remains the same, and the XGBoost algorithm is again used for regression:

$$\hat{y}_i = f(X_i)$$

Where:

- $\hat{y}_i$ is the predicted adjusted closing price for the stock on day i in December 2024.
- X_i includes the same features as the November model, but with data up to November 30, 2024.

The December model is trained using the `XGBoost` of Chen and Guestrin (2024) library and the same parameters as the November model. Once trained, the model will generate predictions for December 2024, leveraging the latest market conditions reflected in the training data.

The assumptions and model design are consistent with the November model, with the key differences being the updated data and the forecast period. By training separate models for November and December, we can capture any evolving trends or shifts in stock price behavior as we approach the end of 2024.

3.4 Model Justification

The XGBoost model is the most suitable approach for forecasting FAANG stock prices in November and December 2024. Our primary outcome variable is the adjusted closing price of each stock, which is continuous and unbounded. XGBoost, a gradient boosting algorithm, excels in handling regression tasks with large datasets, capturing complex, non-linear relationships between features such as lagged prices, moving averages, volatility, and daily returns. To model these relationships effectively, XGBoost's ability to handle feature interactions and non-linearities is key. The inclusion of rolling averages and volatility allows the model to account for both short-term fluctuations and longer-term trends in stock prices, which are often influenced by broader market dynamics. Additionally, the model's flexibility helps capture the intricacies of stock price behavior over time, ensuring that past market patterns are reflected in future predictions. Details are further discussed in **?@sec-model-justification**

While simpler models like linear regression could be used, they would fail to capture the complexity and non-linearity inherent in financial data. Machine learning techniques like random

forests or support vector machines could also provide robust predictions, but XGBoost's combination of accuracy, speed, and interpretability makes it the best choice for this analysis. Ultimately, XGBoost's ability to learn from the data and adjust to market conditions enhances its suitability for predicting FAANG stock prices in an uncertain and volatile market environment.

3.5 Model summary

Both models were evaluated using RMSE, MAE, and R^2 metrics, as shown below in Figure 2. The November model achieved an RMSE of 40.50, MAE of 16.40, and R^2 of 0.988, indicating it captured meaningful variability in the stock price movements. The December model reported significantly lower RMSE (1.10) and MAE (0.78), with an R^2 of 0.999, suggesting near-perfect alignment with the data. However, the December model's near-flat projections likely stem from minimal variability in the placeholder data used for December, resulting in artificially inflated performance metrics.

Metric	November	December
RMSE	40.5042812	1.1038338
MAE	16.4048267	0.7786735
R^2	0.9880415	0.9999909

Figure 2: November and December Model Summary

These results highlight the November model's relative robustness in capturing dynamic trends compared to the December model's overfitted nature due to static input data. This emphasizes the importance of data variability in evaluating predictive model performance.

4 Results

Using the models trained on historical data, we implemented the XGBoost algorithm to predict the adjusted closing prices for each of the FAANG stocks during these two months. For November, since we have the actual adjusted closing prices for November, we compare the predicted values with the actual data to evaluate model performance. For December, since the actual prices for this month are unavailable, the results for December focus solely on the predicted values.
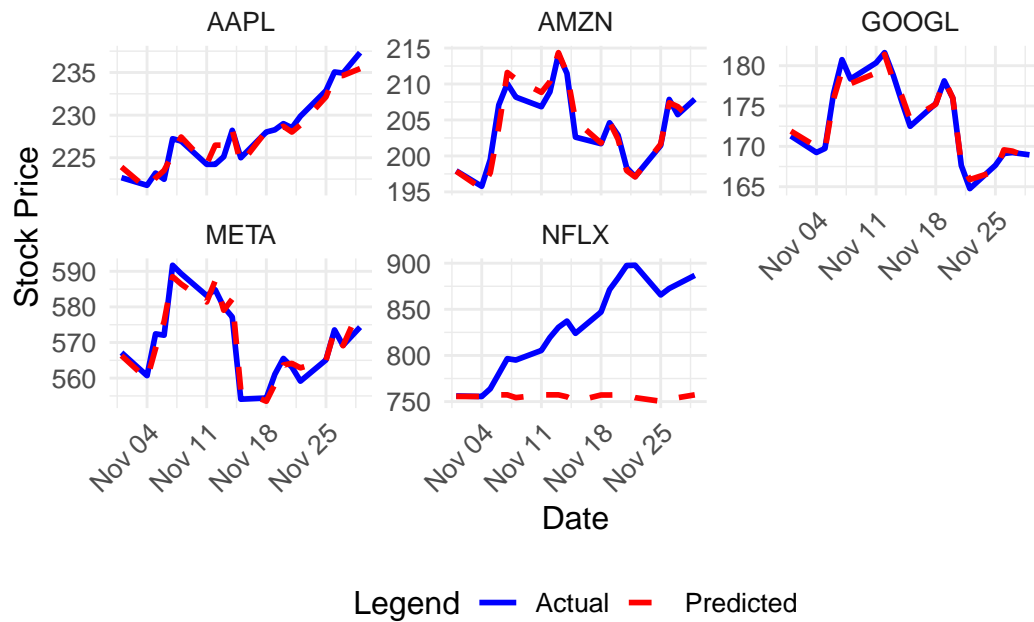
Figure 3: Comparison of actual (blue) vs. predicted (red) adjusted closing stock prices for FAANG companies in November 2024. Highlights the model's ability to capture general trends but shows variability in performance across stocks, particularly for Netflix (NFLX)

4.1 Actual vs Predicted for November 2024

Figure 3 compares actual (blue) and predicted (red) stock prices for FAANG companies during November 2024. The model captures trends effectively for AAPL and META, where predictions align closely with observed prices. However, for NFLX, discrepancies appear later in the month, indicating the model's limitations in capturing certain market volatilities.
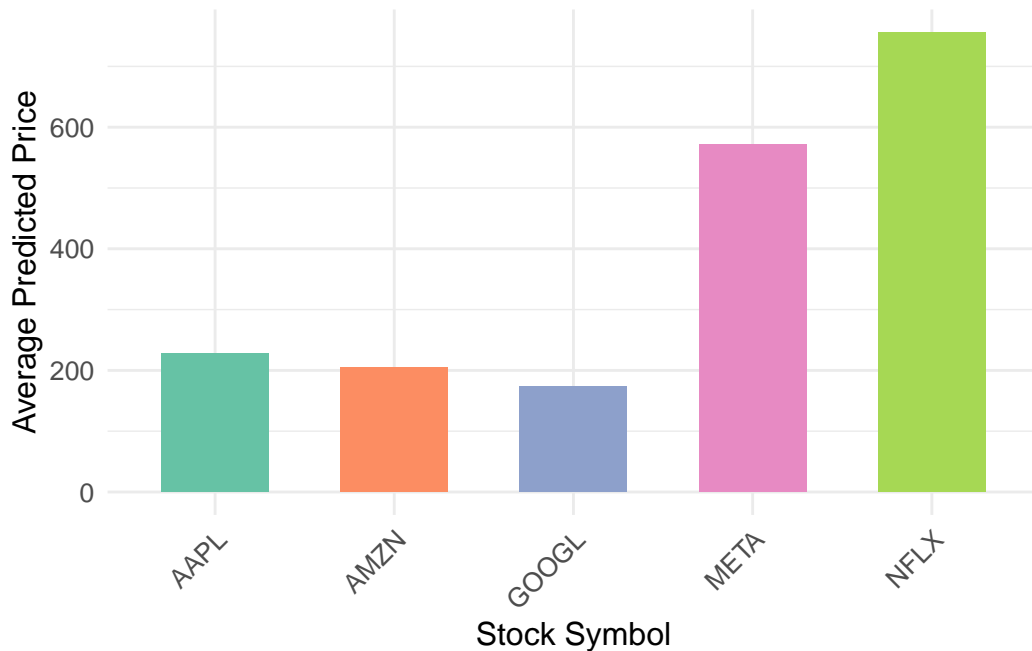
4.2 Average Predicted Stock Price - November

Figure 4: Average predicted stock prices for FAANG companies in November 2024. Netflix leads with the highest predicted average price (~$750), while Google has the lowest (~$180), reflecting varied performance across companies based on model predictions.

Figure 4 shows the average predicted stock prices for FAANG companies in November 2024. Netflix (NFLX) has the highest predicted average price (~$750), followed by Meta (META) at just above $500. Apple (AAPL), Amazon (AMZN), and Google (GOOGL) exhibit lower averages, at ~$225, ~$210, and ~$180, respectively. The chart highlights Netflix's expected strong performance compared to other FAANG companies.
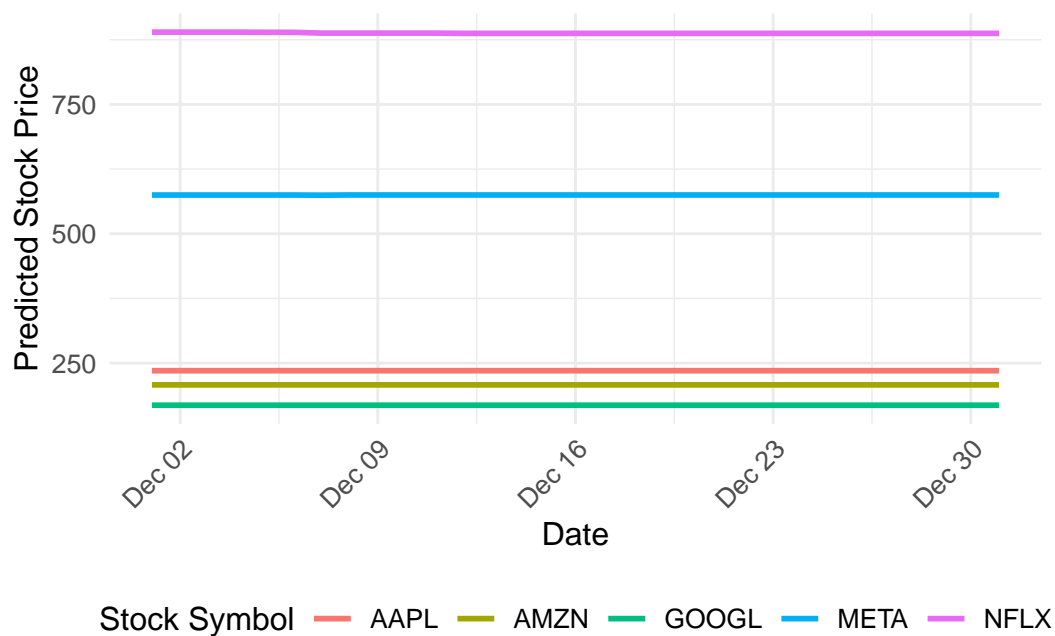
Figure 5: Predicted stock prices for FAANG companies in December 2024. The model predicts relatively flat price trends for all companies, with Netflix having the highest predicted price and Google the lowest

4.3 Predicted Stock Price Trends - December

Figure 5 presents the predicted stock prices for FAANG companies in December 2024, showing flat trends with no significant fluctuations. Netflix (NFLX) has the highest predicted price, while Google (GOOGL) has the lowest. Such flat trends highlight potential limitations in the model's ability to capture real-world stock price volatility, further detailed in **?@sec-model-limitations**.

4.4 Average Stock Prices Bar Chart - December
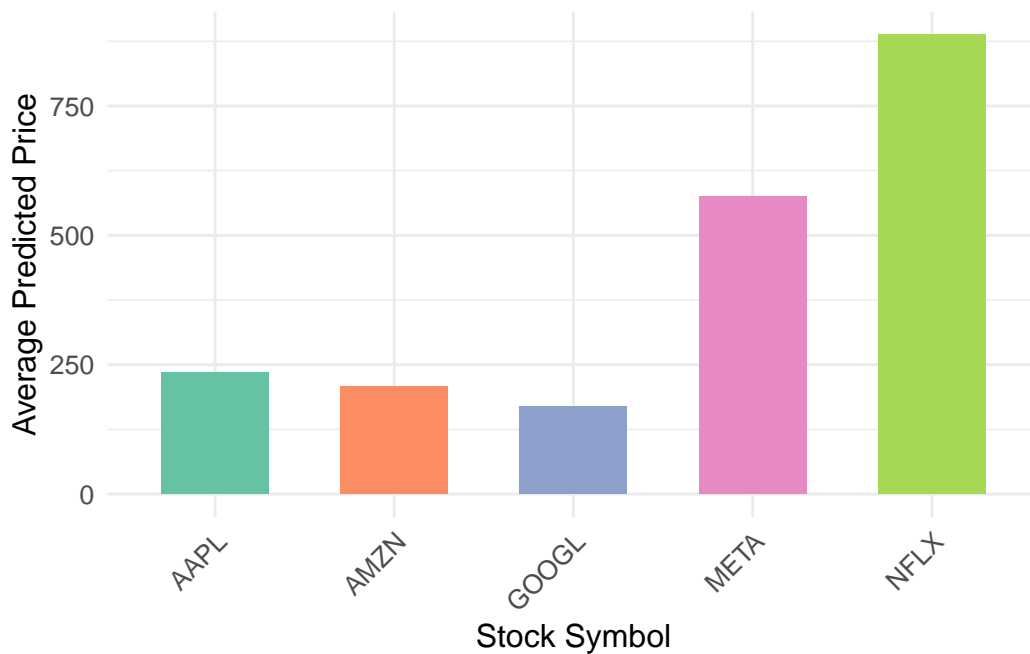
Figure 6: Average predicted stock prices for FAANG companies in December 2024. Netflix leads with the highest predicted price ($825), followed by Meta ($600), while Apple, Amazon, and Google show considerably lower averages.

Figure 6 shows the average predicted December 2024 stock prices for FAANG companies. Netflix (NFLX) leads with the highest predicted price, just around $825, followed by Meta (META) at $600. Apple (AAPL), Amazon (AMZN), and Google (GOOGL) are projected significantly lower, around $250, $200, and $200, respectively. These predictions highlight Netflix's strong expected performance compared to its peers, reflecting trends modeled from historical data.
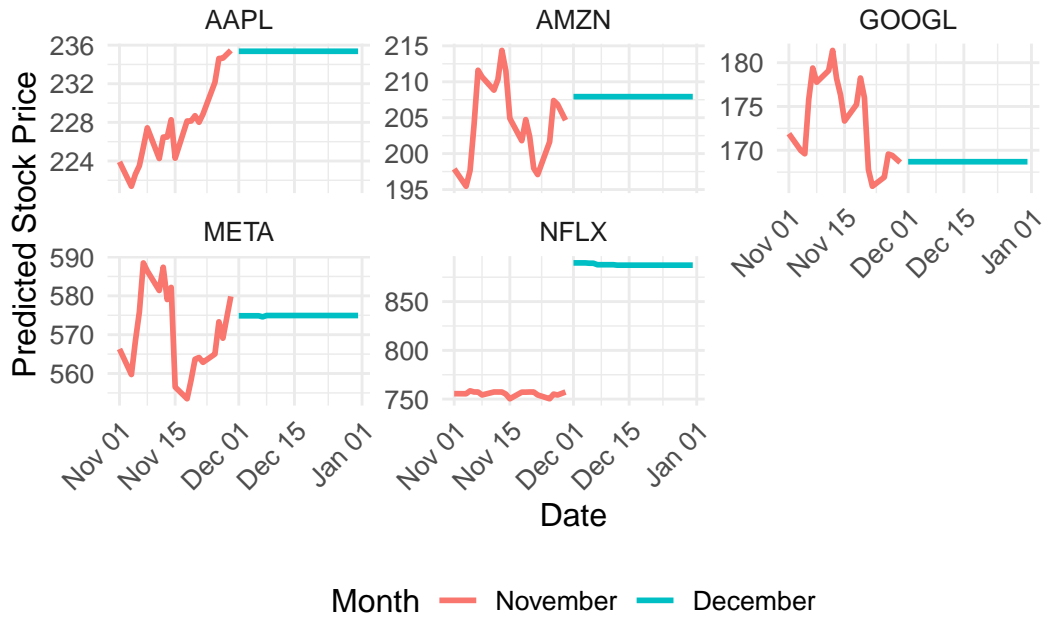
4.5 Combined November and December

Figure 7: Predicted stock prices for FAANG companies across November (red) and December (teal) 2024. November predictions exhibit variability, particularly for Apple (AAPL), Meta (META), Amazon (AMZN), and Google (GOOGL), while December predictions flatten, showing minimal fluctuations. Netflix (NFLX) remains nearly flat across both months. Highlights the model's reduced sensitivity to potential market dynamics in December.

Figure 7 illustrates predicted FAANG stock prices for November (teal) and December (red) 2024. November shows dynamic trends for Apple (AAPL), Meta (META), Amazon (AMZN), and Google (GOOGL), while December predictions flatten across all companies, particularly Netflix (NFLX). This highlights the model's limitation in capturing market volatility and end-of-year dynamics.

4.6 Average Predicted Stock Prices for November and December 2024

Figure 8 compares the average predicted stock prices for FAANG companies in November (red) and December (teal) 2024. Netflix (NFLX) shows the highest predicted prices and a notable increase from November to December. Meta (META) also exhibits a slight increase, while Apple (AAPL), Amazon (AMZN), and Google (GOOGL) show minimal variations, indicating relative stability in their predicted prices across the two months. Only Google
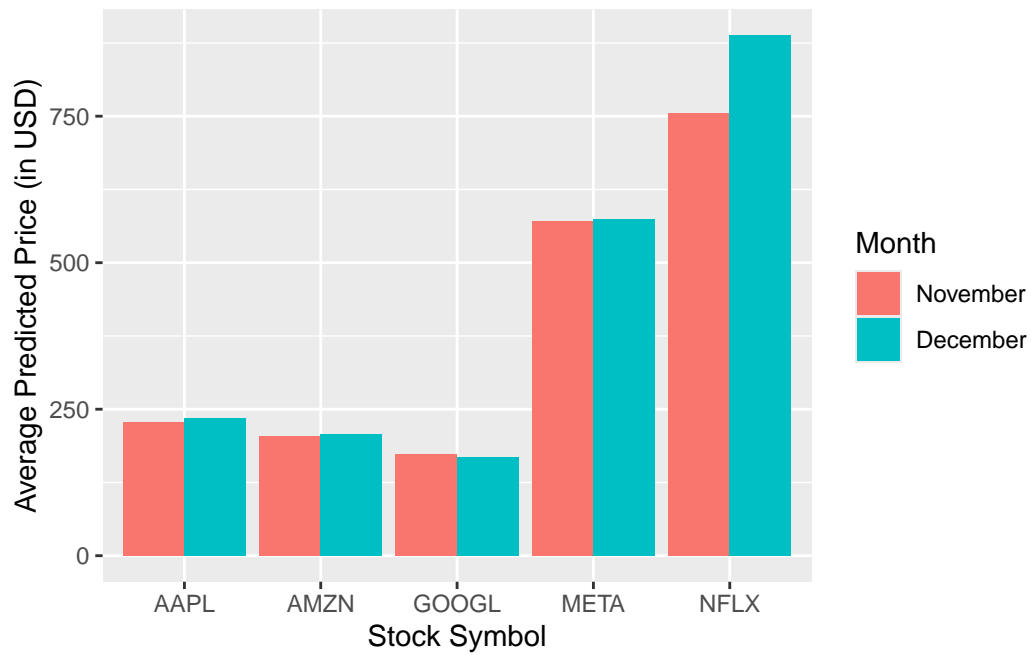
Figure 8: Average predicted stock prices for FAANG companies in November (red) and December (teal) 2024. Netflix shows the highest increase between months, while Apple, Amazon, and Google exhibit stable predicted prices. Only google displays month-to-month decrease.

(GOOGL) displays a slight month-to-month decrease, suggesting for further validation and data analysis.

5 Discussion

5.1 Summary of Findings

The predicted FAANG stock prices for November and December 2024 highlight several trends with real-world implications. In November, the predicted prices closely align with actual stock prices for AAPL, AMZN, and GOOGL, suggesting consistent market behavior across these stocks during the month. However, NFLX predictions were significantly lower than actual prices, likely reflecting unique market dynamics specific to the stock. These dynamics could include sector-specific factors, such as increased investor interest in streaming services, or company-specific announcements, such as major content releases or subscriber growth milestones, that were not captured in the historical data. Additionally, NFLX may be more sensitive to post-election market sentiment compared to other FAANG stocks, contributing to its divergence. Further analysis is needed to fully understand the market behavior of NFLX, particularly in response to macroeconomic and company-specific events during the period.

For December, the steady slope in predicted prices across all stocks, with NFLX and META leading in average value, suggests an expectation of stability or consistent growth. This could reflect broader investor confidence post-election. However, this trend might also stem from the analysis data extending only until November, limiting the model's ability to capture variations within December. These observations suggest that while the predictions provide actionable insights for short-term planning, unique stock-specific dynamics and the influence of incomplete datasets warrant further investigation to fully understand long-term trends.

5.2 Model Performance and Limitations

5.2.1 November Model

The November model demonstrated strong performance in predicting short-term stock price trends. Predictions for AAPL and GOOGL closely aligned with actual prices, indicating the model's ability to effectively leverage historical data and features like moving averages and lagged prices. This suggests that the model captured stable market behaviors for these stocks during November's relatively steady market conditions. However, the model significantly underestimated NFLX prices, pointing to its inability to account for stock-specific dynamics such as sector-specific investor sentiment or company-specific announcements. This limitation arises from the reliance on quantitative features and the exclusion of qualitative factors, such as news events or broader macroeconomic trends. The model's dependence on historical data

also makes it less responsive to sudden or emerging trends, which are particularly impactful for stocks like NFLX.

5.2.2 December Model

The December model's predictions revealed a flat trajectory across all stocks, highlighting its inability to capture intra-month variations and post-election volatility. This limitation is likely due to the analysis data only extending through November, restricting the model's ability to learn patterns or fluctuations specific to December. Even with complete, constructed December data, the reliance on technical indicators and short-term historical trends may have limited the model's ability to reflect dynamic market movements. Stocks like NFLX and META, which showed consistently higher predicted averages, may reflect overfitting to earlier trends, preventing the model from adapting to potential changes in market conditions.

Further model diagnostics, including feature importance evaluations and residual analysis, are provided in the Appendix B.2.

5.3 US Election Effects

Table 3 compares FAANG stock prices one trading day before and after the elections in 2016, 2020, and 2024. Post-election periods show noticeable fluctuations, with some stocks rising and others declining. For instance, AAPL experienced slight increases in 2016 and 2020, while AMZN and NFLX declined in 2016. In contrast, 2024 shows an overall upward trend across all FAANG stocks, with NFLX seeing the largest gain, rising from $755.51 to $780.21.

Table 3: FAANG Stock Prices for One Trading Day Before and After the Election (2016, 2020, and 2024)

year	symbol	Before Election	After Election	Price Change
2016	AAPL	25.56	25.67	↑
2016	AMZN	39.25	38.59	↓
2016	GOOGL	40.00	40.18	↑
2016	META	121.78	122.81	↑
2016	NFLX	124.58	122.19	↓
2020	AAPL	106.18	112.21	↑
2020	AMZN	150.22	162.06	↑
2020	GOOGL	81.02	87.08	↑
2020	META	260.58	286.52	↑
2020	NFLX	484.12	496.95	↑
2024	AAPL	221.77	222.48	↑
2024	AMZN	195.78	207.09	↑

Table 3: FAANG Stock Prices for One Trading Day Before and After the Election (2016, 2020, and 2024)

year	symbol	Before Election	After Election	Price Change
2024	GOOGL	169.24	176.51	↑
2024	META	560.68	572.05	↑
2024	NFLX	755.51	780.21	↑

These trends suggest that market reactions to election outcomes significantly influenced stock prices, particularly in late 2024. This surge likely contributed to the increased stock price predictions for December in our model, which incorporates historical data and trends. The impact of market sentiment and external factors, such as post-election dynamics, highlights the need to consider broader influences when interpreting the model's results.

Appendix

A Additional Data Details

A.1 Data Cleaning

The data cleaning process transformed raw FAANG stock price data from Yahoo Finance into a structured format suitable for predictive modeling. The dataset spans daily stock prices for Meta (META), Amazon (AMZN), Apple (AAPL), Netflix (NFLX), and Alphabet (GOOGL) from January 2020 to November 2024. To prepare for December 2024 predictions, placeholder dates were added to create a continuous sequence of daily observations, accounting for non-trading days such as weekends and holidays. Missing values in the adjusted closing price column were addressed through forward and backward interpolation, ensuring data continuity for subsequent analyses. Rigorous testing ensured no critical missing values remain, all values fall within logical ranges, and the dataset maintain consistency.

A.2 Feature Engineerring

Feature engineering enriched the dataset with additional predictors to improve model performance. Key features included Lag_1 (adjusted closing price from the previous day), Rolling_Mean_7 (7-day rolling average), sma_20 and sma_50 (20-day and 50-day simple moving averages), volatility (20-day rolling standard deviation), and daily_return (percentage change in adjusted closing price). These features captured temporal trends, smoothed short-term fluctuations, and quantified variability in stock prices. Stock symbols were numerically encoded to help the model identify patterns across different stocks. These features are further explained in Appendix B.1.2

A.3 Data Limitations and Future Directions

While the cleaned dataset is robust and suitable for modeling, several limitations remain. External factors, such as macroeconomic indicators, earnings announcements, or industry-specific news, are not included, despite their potential to significantly impact stock prices. The interpolation of missing values assumes smooth price continuity, which may not fully account for abrupt market changes or high-impact events occurring during non-trading days. Additionally, the analysis is restricted to FAANG stocks, which, although influential, may not capture broader market dynamics or indirect effects from other sectors. These limitations may reduce the model's ability to fully represent the complexities of the market.

A.4 Furutre Directions

Future analyses could address these limitations by incorporating additional data sources, such as macroeconomic indicators or sector-specific variables, to enhance predictive accuracy. Expanding the dataset to include other major stocks or sectors could provide a more comprehensive view of market behavior and its impact on FAANG stocks. Advanced feature engineering techniques, such as seasonal decomposition, non-linear transformations, or incorporating time-varying covariates, could improve the models' ability to capture complex interactions and nuanced market behaviors. These improvements would strengthen the robustness and generalizability of the analysis, offering deeper insights into both short- and long-term trends in FAANG stock dynamics.

B Additional Model Details

B.1 Model Justification - Further Details

The XGBoost regression model was chosen for predicting FAANG stock prices in December 2024 due to its ability to model complex non-linear relationships, which are common in financial time series data. Given the dynamic and time-sensitive nature of stock prices, the November and December models were trained separately to ensure that predictions account for evolving market trends and feature importance within each specific period.

B.1.1 Why Separate November and December Models?

Separate models for November and December were developed to capture the unique characteristics of stock price movements during these months. By training the November model using data up to October 31, 2024, and the December model using data up to November 30, 2024, we ensure that each model incorporates the most recent market information available for its respective forecasting period. This separation minimizes over-fitting to earlier trends and allows the December model to adapt to any shifts in market behavior or volatility observed in November.

B.1.2 Model Features

XGBoost was selected due to its ability to handle non-linear interactions and temporal dependencies, which are crucial for stock price prediction. Key engineered features include:

- **Lag_1**: Represents the adjusted closing price from the previous day, capturing short-term price dependency.

- **Rolling_Mean_7**, **sma_20**, **sma_50**: Reflect short- and medium-term trends, offering insights into market behavior over recent periods.
- **Volatility**: Captures price variability using the 20-day rolling standard deviation, a critical factor in understanding market risk.
- **Daily_Return**: Quantifies day-to-day price changes, highlighting rate-of-change trends.
- **Symbol_Encoded**: Encodes stock-specific identifiers numerically, enabling the model to generalize patterns across multiple FAANG stocks.

By incorporating these features, the model is able to account for short-term fluctuations and medium-term market dynamics, which are critical for accurate stock price forecasting.

B.1.3 Hyperparameters and Regularization

The November and December models share key hyperparameters optimized to balance complexity and generalizability: - **Max depth**: Set to 6, allowing the model to capture significant interactions without overfitting. - **Learning rate**: Set to 0.1, ensuring gradual refinement of predictions. - **Boosting rounds**: Fixed at 100 to provide sufficient iterations for model convergence. XGBoost's built-in regularization techniques mitigate overfitting, especially important given the high-dimensional engineered features and the presence of missing values due to holidays and weekends.

Unlike traditional time-series models such as ARIMA, which assume stationarity and often require extensive preprocessing, XGBoost directly handles non-stationary data. This flexibility makes it an ideal choice for stock price forecasting, where trends and volatility can vary significantly over time. Furthermore, XGBoost automatically handles missing data, making it robust for real-world applications with gaps in trading days. The decision to train separate models for November and December, coupled with XGBoost's ability to handle complex non-linearities, ensures that the forecasts reflect both temporal dependencies and evolving market dynamics. By balancing model flexibility and interpretability, the November and December XGBoost regression models provide a robust and reliable framework for predicting FAANG stock prices, tailored to the unique characteristics of each time period.

B.2 Model Diagnostics

B.2.1 Feature Importance

Figure 9 for the November model reveals that Lag_1 (the previous day's adjusted closing price) is the most influential predictor, accounting for the majority of the model's decision-making. Rolling_Mean_7 (the 7-day rolling average) is the second most significant feature, albeit with a much smaller contribution. Other features, such as daily_return, sma_20, and volatility, show negligible impact. This indicates that the model primarily relies on short-term

trends and immediate past values to predict stock prices, reflecting its focus on short-term forecasting.
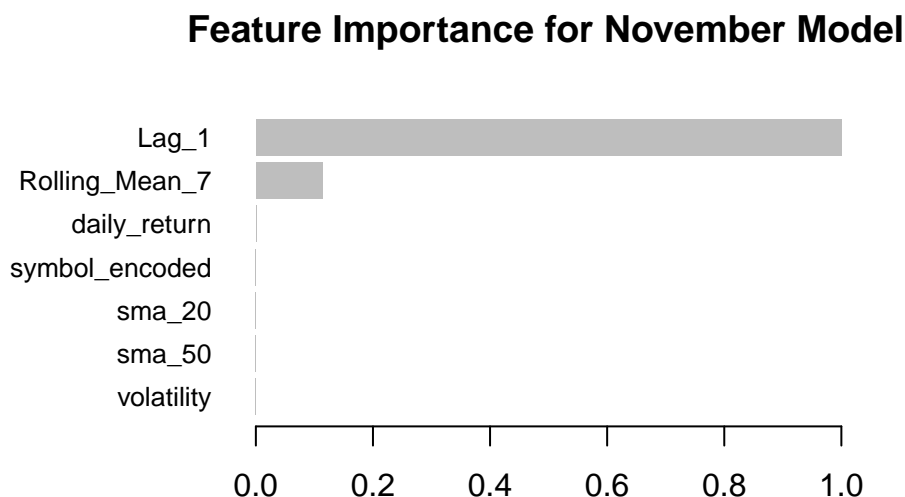
Feature Importance for November Model

Figure 9: November Feature Importance

B.2.1.1 December Model

Figure 10 reveals that for the December model, Lag_1 remains the dominant feature, followed by Rolling_Mean_7, which has a slightly larger contribution compared to the November model. The remaining features, including daily_return, sma_20, and volatility, exhibit minimal importance, similar to the November model. The heavy reliance on short-term features like Lag_1 and Rolling_Mean_7 underscores the model's emphasis on recent price trends for prediction, while broader, longer-term patterns play a negligible role in its forecasting approach.

B.2.2 Residual Analysis

B.2.2.1 November Model

Figure 11 evaluates the November model's performance by analyzing the residuals (errors) against the predicted (fitted) values. The residuals vs. fitted values plot for the November model shows that residuals are evenly distributed around zero for most fitted values, suggesting that the model captures the central trend of the data well. However, there are slight patterns
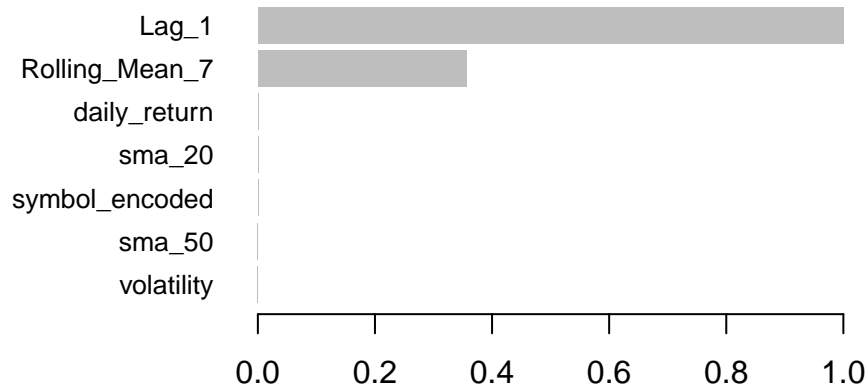
Feature Importance for December Model

Figure 10: Feature Importance for December Model

in residuals at higher fitted values, indicating some systematic deviations that the model may not fully account for. The majority of residuals fall within a narrow range, suggesting good model accuracy overall, but some outliers indicate the potential influence of unmodeled factors or noise in the data.

B.2.2.2 December Model

Figure 12 reveals a similar pattern to the November model, with residuals mostly centered around zero. However, the residuals exhibit slightly tighter clustering, reflecting the December model's reliance on flatter trends in the data. Outliers are still present, but their distribution is more uniform across fitted values. This tighter residual range suggests the December model captures the data's central trend but is less sensitive to short-term volatility, as reflected in the flat December predictions.

B.2.3 Error Distribution

B.2.3.1 November Model

Figure 13 reveals that error distribution for the November model is approximately normal, with residuals centered around zero and a slight skewness at the tails. Most residuals are
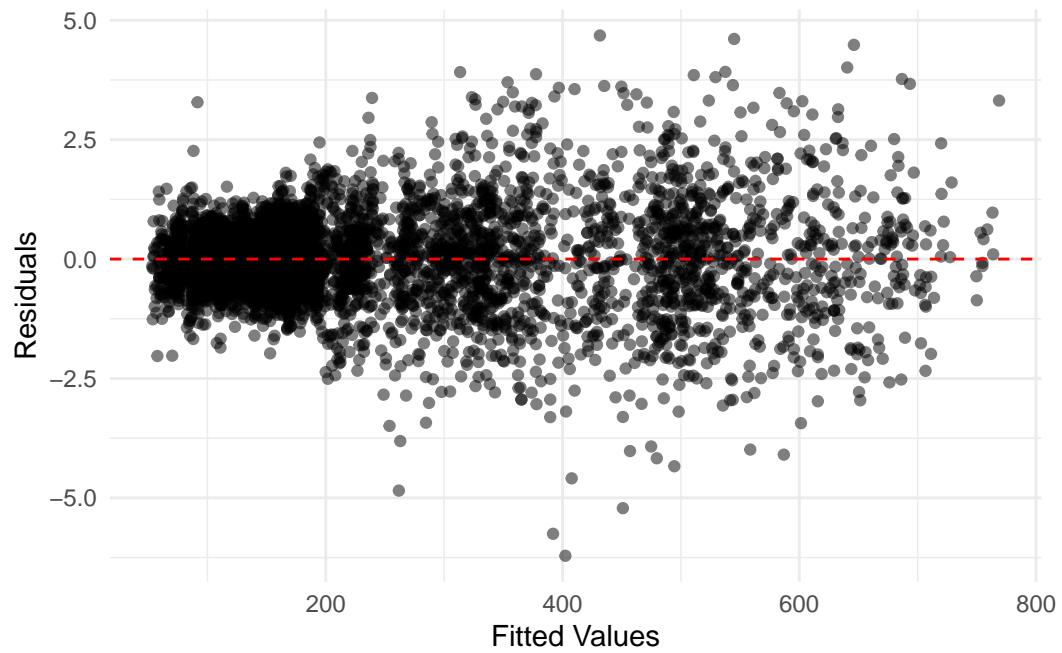
Figure 11: Residuals vs. fitted values for the November model. The residuals are centered around zero, indicating accurate predictions overall, though some systematic deviations and outliers are observed at higher fitted values.
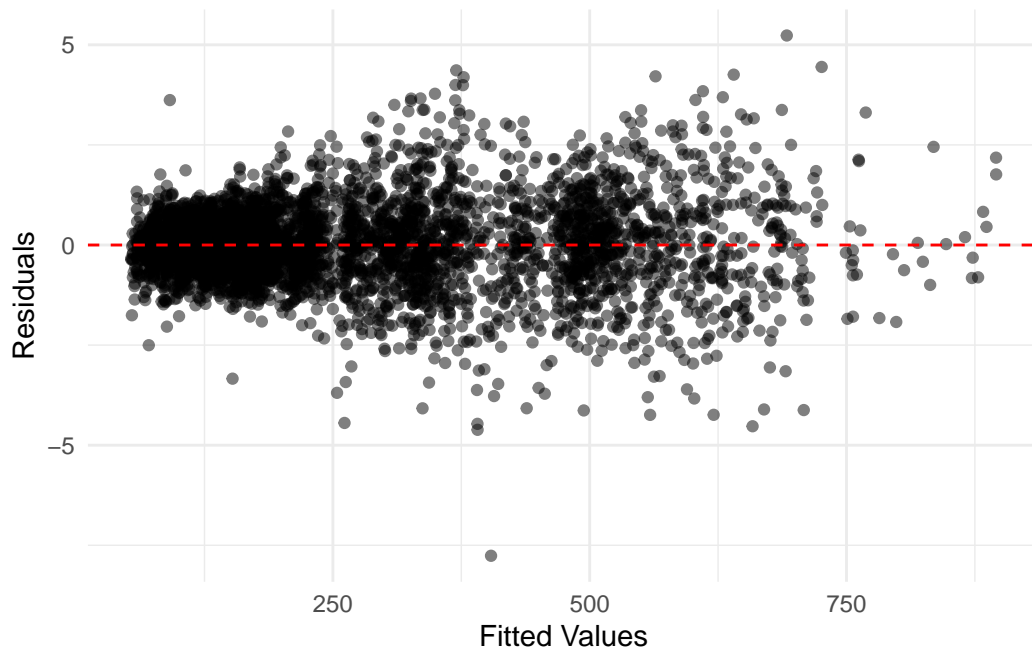
Figure 12: Residuals vs. fitted values for the December model. The residuals are more tightly clustered around zero, reflecting the model's reliance on flatter trends but with less sensitivity to short-term volatility.
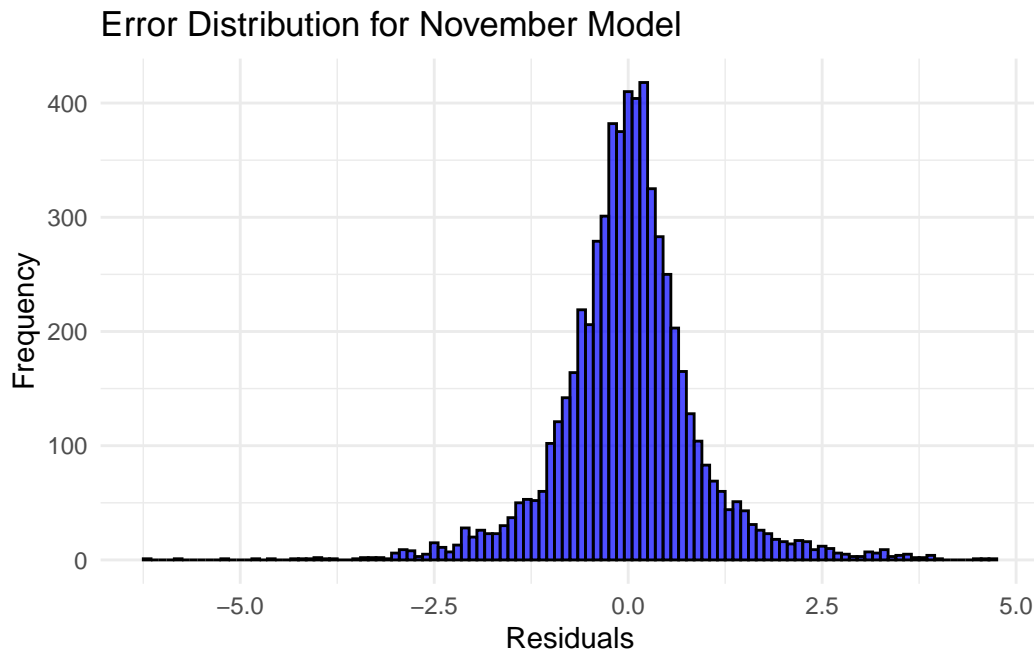
Figure 13: Histogram of residuals for the November model. Residuals are centered around zero with most values falling within -2.5 to 2.5, reflecting accurate predictions with minimal systematic error.

concentrated within the range of -2.5 to 2.5, indicating the model predicts stock prices with a high degree of accuracy. However, the presence of a few outliers suggests the model struggles with certain extreme price movements.
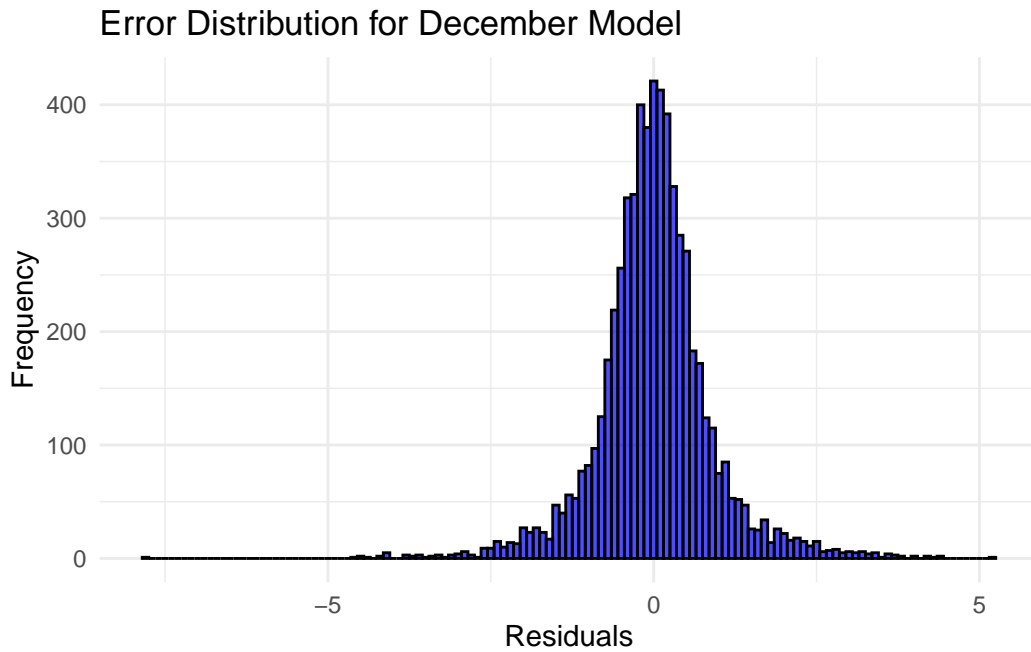

Figure 14: Histogram of residuals for the December model. Residuals are tightly centered around zero, indicating consistent predictions but limited sensitivity to abrupt stock price changes.

?@fig-dec-error-distrib shows an approximate normal error distribution for the December Model, with residuals tightly clustered around zero. Compared to the November model, the distribution is slightly narrower, suggesting reduced variability in prediction errors. This reflects the December model's reliance on flatter trends and its inability to capture sharp price movements, leading to smaller but more consistent errors.

B.2.4 Prediction vs. Actual Values (November)

Figure 15 demonstrates a near-perfect alignment between predicted and actual stock prices, as indicated by the data points closely following the 45-degree reference line. This suggests that the model accurately captures the stock price trends with minimal deviation. The tight clustering of points along the line reflects the model's strong predictive performance, supported by the high R^2 value observed in the evaluation metrics.

Such visualization cannot be replicated for the December model because the predicted values for December are based entirely on placeholder dates, with no actual stock price data available
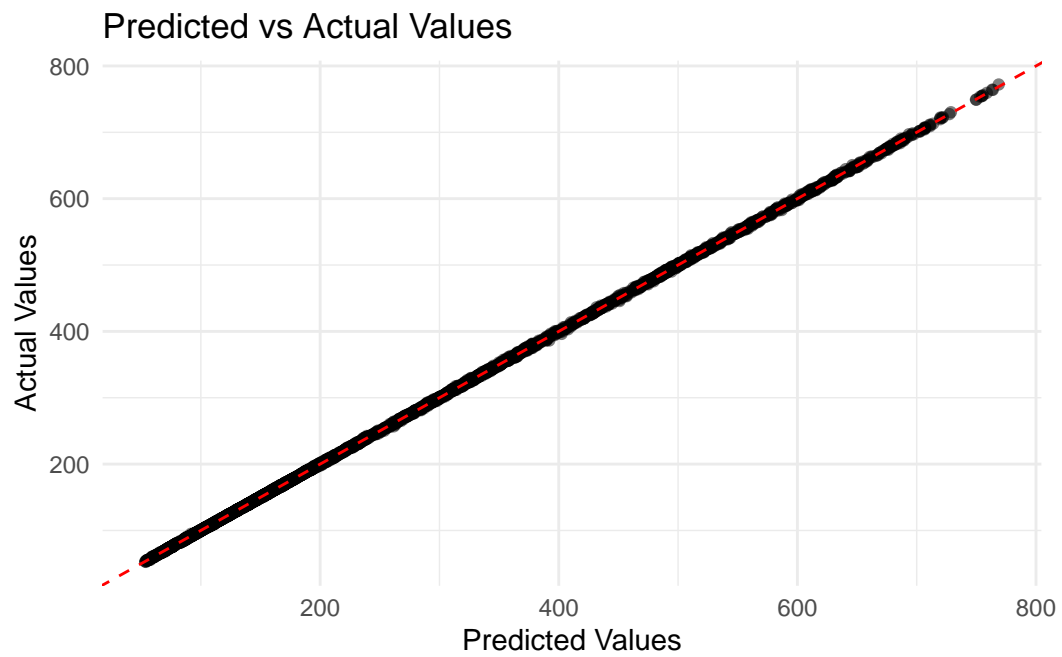
Figure 15: Predicted vs. actual stock prices for the November model. The points closely follow the 45-degree line, indicating strong predictive accuracy and alignment with actual values.

for comparison. As such, there are no actual values for December against which to evaluate the predictions, making it impossible to construct a meaningful predicted vs. actual plot.

B.3 Model Performance and Limitations

B.3.1 November Model Performance

The November model demonstrates strong predictive accuracy, as shown by the close alignment between predicted and actual values along the 45-degree reference line. Residuals are centered around zero and follow an approximately normal distribution, with most residuals falling within the range of -2.5 to 2.5. These results indicate that the model effectively captures short-term trends and variability in FAANG stock prices. Feature importance analysis highlights Lag_1 (previous day's adjusted price) and Rolling_Mean_7 (7-day rolling average) as the most influential predictors, confirming the model's ability to leverage recent price dynamics for accurate forecasting. However, the model exhibits limitations in capturing extreme price movements or unusual events not reflected in the historical data. Slight patterns in residuals at higher fitted values and outliers in the error distribution suggest that the model struggles to account for high-impact or sector-specific factors, limiting its generalizability to unexpected market conditions.

B.3.2 December Model Performance

The December model, trained on data up to November 30, 2024, shows significant limitations in its predictive capabilities. Residuals are tightly clustered around zero, with a narrower error distribution compared to the November model. While this suggests minimal noise in predictions, it reflects an overreliance on flat trends, resulting in an inability to capture dynamic stock price movements. The flat slopes in predicted values, combined with the absence of actual December stock price data for evaluation, underscore the model's failure to incorporate potential market shifts, such as post-election sentiment or end-of-year trading dynamics. Similar to the November model, feature importance analysis identifies Lag_1 and Rolling_Mean_7 as dominant predictors. However, the December model's heavy dependence on these features exacerbates its inability to account for broader market influences, leading to overly static predictions.

B.3.3 Model Limitations

While the November and December XGBoost models demonstrate strong performance, several limitations should be noted:

1. **Overreliance on Short-Term Features**: Both models heavily depend on Lag_1 and Rolling_Mean_7, which focus on short-term trends. While effective for stable markets, this reliance limits their ability to predict abrupt changes or long-term patterns influenced by broader macroeconomic or sector-specific factors.

2. **Exclusion of External Influences**: Neither model incorporates external variables such as macroeconomic indicators, earnings reports, geopolitical events, or industry-specific news, which are critical in driving stock price movements. This omission reduces the models' capacity to adapt to dynamic market conditions.

3. **Flat Predictions for Certain Periods**: The flat predictions observed in the December model highlight a broader limitation: both models assume continuity in trends and fail to account for potential volatility or sudden market shocks. This may be a result of insufficient diversity in the feature set.

4. **Lack of Cross-Validation Across Time Periods**: While each model is trained separately on its respective time frame, there is no mechanism to ensure their consistency across overlapping or adjacent periods. This could lead to discrepancies when transitioning from November to December predictions.

5. **Sensitivity to Data Availability**: The models rely entirely on historical stock price data, meaning any gaps or interpolation inaccuracies could propagate into the predictions. This limitation is particularly relevant for the December model, where placeholder dates and interpolated values were used for prediction.

Addressing these limitations in future analysis, such as incorporating external data and improving the feature set to capture dynamic market behaviors, would enhance the robustness and applicability of the models.

C Surveys, Sampling, and Observational Data

C.1 Nature of Observational Data

The stock price data used in this study is observational, derived from naturally occurring financial transactions in the stock market. Unlike experimental or survey data, observational data reflects real-world behaviors influenced by external factors such as macroeconomic policies, geopolitical events, and investor sentiment. This complexity makes it challenging to distinguish causality from correlation, as unmeasured confounding variables can bias interpretations. For example, stock price movements during the 2024 U.S. Presidential Election may reflect both direct market reactions and broader economic conditions.

While observational data offers a granular, high-frequency view of stock market dynamics, it relies entirely on historical trends as proxies for predictive features. This approach limits its ability to account for unprecedented shocks, such as global crises or election-related volatility,

which may diverge from historical patterns. Furthermore, temporal dependence—where current prices are influenced by past values—requires careful feature engineering, such as lagged variables (Lag_1) and rolling averages (Rolling_Mean_7), to capture meaningful trends.

C.2 Sampling Frame and Representativeness

The sampling frame encompasses all trading days from January 2020 to November 2024, providing a comprehensive record of FAANG stock activity during a period of significant market and political events. While this dataset ensures detailed coverage of the technology sector, it is constrained by its sector-specific focus, which limits generalizability to other industries or broader market trends.

C.2.1 Strengths and Weaknesses

A key strength of the dataset is its high-frequency nature, which allows the analysis to capture both short-term fluctuations and long-term trends. FAANG stocks, as market leaders, serve as valuable indicators of overall economic conditions and investor sentiment. However, the focus on a single sector introduces limitations in generalizability, as trends within technology stocks may not reflect behaviors across other industries. Furthermore, the imputation of placeholder prices for December 2024 ensures continuity but risks underestimating volatility, particularly in the context of post-election market adjustments.

C.3 Integrating Surveys with Observational Data

Surveys provide forward-looking indicators that enhance the interpretation of stock price trends. For example, during significant events like the 2024 U.S. Presidential Election, surveys could capture investor sentiment regarding anticipated market impacts, which may not yet be reflected in observed price movements.

C.3.1 Proposed Survey Design

To complement the observational stock data, investor sentiment surveys could include questions such as:

- **Market Confidence**: "Do you expect FAANG stock prices to increase, decrease, or remain stable over the next month?"
- **Election Impact**: "How has the 2024 U.S. Presidential Election influenced your investment decisions?"
- **Risk Appetite**: "What is your likelihood of investing in technology stocks in the current market climate?"

C.3.2 Methodological Considerations:

- **Sampling and Representativeness**: Stratify respondents by demographics, investment experience, and portfolio size to ensure diverse perspectives.
- **Weighting**: Apply post-survey weighting to align survey demographics with actual investor populations, similar to post-stratification in traditional surveys.

C.3.3 Strengths and Limitations

Advantages of Survey Integration:

- **Contextualization**: Surveys can help interpret stock price trends, particularly around major events like elections.
- **Predictive Power**: Sentiment data may improve model accuracy by capturing market expectations and behavioral drivers.

Limitations of Survey Integration:

- **Response Bias**: Survey participants may exaggerate optimism or pessimism, especially during volatile periods.
- **Sampling Complexity**: Ensuring representativeness across demographics, such as institutional vs. retail investors, requires significant resources and methodological rigor

C.4 Parallels Between Observational Data and Surveys

Observational stock data and surveys share structural similarities that make their integration intuitive and mutually reinforcing:

- **Stratification**: The division of stock data by symbol mirrors stratified sampling in surveys, allowing for detailed analysis at the stock level while preserving aggregate insights.
- **Imputation**: Both surveys and stock data require robust imputation methods to address missing values. For example, placeholder stock prices for December 2024 were interpolated using forward and backward filling, similar to imputing missing survey responses.
- **Weighting**: Models could incorporate macroeconomic indicators, such as GDP or interest rates, as weights to adjust for external market influences, akin to post-stratification in surveys.

C.5 Addressing Data Challenges

- **Temporal Dependence**: Stock prices are influenced by past trends. To mitigate this, features such as lagged prices (Lag_1) and rolling averages (Rolling_Mean_7) were used to capture momentum and smoothed trends effectively. However, these features may miss sudden shifts caused by external shocks or policy changes.

- **Imputation for Missing Data**: Missing December 2024 values were filled through interpolation to ensure continuity. While necessary for analysis, this approach may underestimate volatility or fail to account for post-election market adjustments. Advanced time-series models could improve the accuracy of such imputations.

- **Observational Bias**: Stock data is inherently influenced by external confounders, such as Federal Reserve policies, geopolitical tensions, or major corporate announcements. These confounders obscure causal relationships. Future models could include external predictors to better account for such influences, improving robustness.

C.6 Recommendations for Future Work

1. **Survey Integration**: Combine investor sentiment surveys with stock data to capture qualitative dimensions of market behavior. Surveys can reveal forward-looking insights, such as investor confidence, that improve interpretability and predictive power.
2. **Expanding Scope**: Extend the analysis to include sectors beyond FAANG stocks to enhance generalizability and uncover trends across diverse industries.
3. **Improved Imputation**: Use dynamic time-series models to address missing values more accurately and capture volatility during critical events like elections or earnings announcements.

By addressing these challenges and integrating surveys with observational data, future analyses can achieve a richer, more nuanced understanding of market dynamics. Such hybrid approach provides a comprehensive framework for interpreting stock price trends, making predictions more robust and actionable, especially during periods of significant economic or political change.

References

Chen, Tianqi, and Carlos Guestrin. 2024. *xgboost: Extreme Gradient Boosting.* Distributed Machine Learning Community. https://xgboost.readthedocs.io/.

Dancho, Matt. 2024. *tidyquant: Tidy Quantitative Financial Analysis.* Business Science Corporation. https://business-science.github.io/tidyquant/.

Foundation, Apache Software. 2024. *arrow: Apache Arrow and Parquet Support for R.* https://arrow.apache.org/.

Grolemund, Garrett, and Hadley Wickham. 2024. *lubridate: Make Dealing with Dates a Little Easier.* RStudio, PBC. https://lubridate.tidyverse.org/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Team, Yahoo Finance. 2024. *Yahoo Finance API for Stock Data.* https://finance.yahoo.com/.

Ulrich, Joshua. 2024. *TTR: Technical Trading Rules.* https://CRAN.R-project.org/package=TTR.

Wickham, Hadley, and tidyverse contributors. 2024. *tidyverse: Easily Install and Load the Tidyverse.* RStudio, PBC. https://www.tidyverse.org/.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2024. *dplyr: A Grammar of Data Manipulation.* RStudio, PBC. https://dplyr.tidyverse.org/.