Parsing the Fed: News, Noise and Forecasting Market Returns

Jalen Buffert
April 23, 2025
Gettysburg College
Buffja01@gettysburg.edu

Abstract

Predicting stock prices has perpetual interest among investors, students, researchers, and policymakers. Financial markets are competitive and fast-paced, with all participants seeking an edge over the next. This study evaluates the effectiveness of incorporating sentiment analysis into stock price prediction. Sentiment is extracted from global news sources and FOMC transcripts and merged with historical price data from Yahoo Finance. Natural Language Processing (NLP) is introduced through FinBERT, an extension of the BERT transformer model fine-tuned for financial text. The resulting sentiment values are integrated into ARIMA, SARIMAX, and XGBoost models to evaluate predictive power. The findings reveal strong evidence that neutral sentiment from FOMC communications is inversely related to market returns. While sentiment alone does not consistently yield statistically significant results, combining it with machine learning demonstrates the potential for robust and accurate forecasting.

Honor Code: Jalen Buffert

I affirm that I have upheld the highest principles of honesty and integrity in my academic work and have not witnessed a violation of the Honor Code.

Introduction

In 2025, information circulates across the internet at unprecedented speed, and growing debate surrounds the actual value of that information in financial markets. There has been substantial dialogue in the literature on whether sentiment analysis possesses the predictive power to forecast stock returns. A robust body of research suggests that sentiment derived from social media, financial news, and earnings reports can offer meaningful signals about future price movements. Sentiment, broadly defined, reflects a prevailing attitude or opinion toward a situation or event. Sentiment analysis, in turn, is the systematic extraction of tone or polarity from large volumes of text. In periods of market inefficiency, investor emotion often influences trading behavior, shaping price dynamics in ways not entirely explained by fundamentals. When sentiment becomes widespread, it forms a consensus that often markets price in. This highlights the growing relevance of sentiment analysis; it presents opportunities to generate idiosyncratic alpha by exploiting how markets digest and react to qualitative information.

Quantifying sentiment can be challenging, but a growing body of literature leverages Natural Language Processing (NLP) techniques to extract tone and polarity from unstructured text. NLP is a subset of machine learning that enables computer programs to interpret, understand, and derive meaning from human language. A major breakthrough came with Bidirectional Encoder Representations from Transformers (BERT), introduced by Google in 2018, a pretrained, transformer-based architecture that processes text in both directions, allowing for deeper contextual understanding. However, models like BERT can still struggle with sarcasm, idiomatic expressions, or domain-specific language. In the context of financial analysis, where terminology and phrasing often differ from general text, this limitation is especially important. To address this, the Financial BERT (FinBERT) model was developed as a fine-tuned extension of BERT trained on financial corpora. FinBERT captures nuances in financial communication that traditional NLP models typically miss. In this study, sentiment is extracted using FinBERT and applied to Federal Open Market Committee (FOMC) meeting transcripts, an essential input for modeling how macroeconomic tone influences equity returns.

In recent years, advances in quantitative trading have resulted in roughly 70% of all market transactions being executed algorithmically. In this environment, analyzing public sentiment surrounding publicly traded assets, particularly in the context of policymaker meetings, financial media, and social platforms, has become critical for understanding asset price dynamics and enhancing predictive models. If sentiment analysis produces statistically significant results supporting its ability to forecast asset prices, it raises questions of informational asymmetry. Specifically, do institutional investors leveraging advanced sentiment models gain an unfair edge over retail traders? Moreover, as sentiment-driven strategies grow in popularity, could markets begin reacting more acutely to news events, amplifying volatility, and destabilizing short-term pricing? These concerns intersect with how the Federal Open Market Committee (FOMC) and other policymakers engage in forward guidance. If sentiment measurably influences market outcomes, central banks may need to re-evaluate the clarity, tone, and framing of their communications. A final implication is that high-frequency trading firms and hedge funds may be uniquely positioned to capitalize on these sentiment signals at scale, raising broader questions about market access and fairness.

The Efficient Market Hypothesis (EMH) asserts that all publicly available information is fully reflected in asset prices at any given time. However, its validity has been widely contested in empirical literature. During periods of market inefficiency, sophisticated high-frequency trading (HFT) firms often exploit short-lived mispricings. As these arbitrage opportunities are acted upon, markets tend to revert toward equilibrium. If sentiment analysis demonstrates predictive power, it lends support to the framework of behavioral finance, which emphasizes the role of investor psychology and emotion in driving price movements. This raises important empirical questions: *Does sentiment analysis of macroeconomic news help predict stock market behavior? And can a sentiment-based trading strategy consistently outperform traditional benchmarks?*

This study evaluates the effectiveness of macroeconomic sentiment in predicting stock returns. Five firms (AAPL, BRK.B, MCD, TM, and XOM) were selected across four key sectors: Information Technology, Financials, Consumer Discretionary, and Energy. Market data was sourced from Yahoo Finance, while sentiment data was extracted from the GDELT Project and FOMC meeting transcripts. Additional macroeconomic indicators were retrieved from the Federal Reserve Economic Data (FRED) database. All data sources were aligned and interpolated to match the monthly frequency of stock returns. To assess predictive strength, sentiment-augmented models were compared to a traditional ARIMA baseline. For enhanced robustness and nonlinear pattern capture, an Extreme Gradient Boosting (XGBoost) model was also implemented.

The remainder of this paper proceeds as follows. Section 2 describes the literature and how this study differs from current work. Section 3 describes the processes and methodology taken to make forecasts based on macro sentiment. Section 4 explores the data and explains how we gauged sentiment. Section 5 presents the results of the ARIMA and XGBoost models. Section 6 concludes and discusses future work.

Literature Review

The literature relevant to this research is extensive. This study seeks to explore the relationship between macroeconomic sentiment, sourced from FOMC meeting transcripts, social media platforms, financial news outlets, and online forums, and stock market returns. Numerous studies have established that sentiment extracted from such sources can exhibit predictive power over future price movements. Kaffel et al. (2024), for example, applied advanced Natural Language Processing (NLP) techniques and Large Language Models (LLMs) to financial tweets and uncovered a significant correlation between public sentiment and asset prices. Similarly, Makrehchi et al. (2013) and Oh and Sheng (2011) demonstrated that sentiment signals can serve as strong predictors of directional market shifts. More recently, Panpoonsup et al. (2022) employed FinBERT in conjunction with a Long Short-Term Memory (LSTM) neural network to fuse sentiment analysis with historical pricing data, enhancing the accuracy of stock return forecasting.

Empirical research in behavioral finance suggests that investor sentiment, often observable through retail investor demand, can cause asset prices to diverge from their intrinsic values (Kim & Kim, 2014). Periods of elevated sentiment typically see increased investor exposure to riskier assets, which drives valuations above fundamental levels. Eventually, these mispricings tend to correct as markets revert toward equilibrium. This dynamic supports the findings of Balsara et al. (2008), who showed that contrarian strategies, especially those grounded in technical trading signals, consistently outperformed traditional buy-and-hold benchmarks.

This research also explores sentiment in the context of macroeconomic events, building on the work of Shapiro and Wilson (2021). Their study revealed that sentiment conveyed by policymakers can be leveraged to estimate the Federal Reserve's loss function. By quantifying negative sentiment embedded within internal FOMC discussions, they developed a novel proxy for the Fed's underlying policy objectives, shedding light on how central banks weigh trade-offs when setting monetary policy.

The proposed strategy begins by collecting text data from financial news sources and FOMC transcript archives. These platforms offer web scraping access or APIs, which are queried using Python. Sentiment scores are then extracted using the FinBERT model, which analyzes and assigns sentiment values to each document. As a baseline, an AutoRegressive Integrated Moving Average (ARIMA) model is implemented to capture underlying temporal patterns in returns. To incorporate external sentiment variables, a Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors (SARIMAX) model is introduced to evaluate whether sentiment improves forecast accuracy. To better capture potential nonlinear relationships, the final model applied is Extreme Gradient Boosted Trees (XGBoost), a machine learning algorithm widely used in predictive modeling. Following model development, an event-driven trading strategy is backtested based on sentiment and macroeconomic news responses, and its performance is compared against a buy-and-hold benchmark. Lastly, statistical tests are conducted to evaluate significance, and broader financial and policy implications are discussed.

This study contributes to the existing literature in several key areas. Whereas much of the prior research has concentrated on firm-specific sentiment signals, this analysis shifts the focus toward macroeconomic sentiment, capturing the influence of broader economic events on overall market behavior. Additionally, while most sentiment-based models rely on a single data source, this research incorporates multi-source sentiment aggregation to construct a more robust and representative sentiment measure. Most importantly, if the empirical results demonstrate that sentiment significantly predicts stock returns, the findings will challenge the foundational principles of the Efficient Market Hypothesis (EMH), revealing persistent pricing inefficiencies driven by investor psychology and information interpretation.

Although extensive literature supports the effectiveness of sentiment analysis in predicting market returns, some studies, such as Oliveira (2013), have found no evidence of significant return predictability using sentiment indicators. However, this does not present a limitation to my project. Regardless of whether a significant correlation is observed, understanding the relationship between sentiment and stock returns remains valuable. One potential challenge is acquiring sector-specific data. If this issue arises, I plan to address it in one of two ways: first, by utilizing Interactive Brokers' API and tick data, which would incur a cost but is feasible since I would already be a customer; or second, by using daily stock data, which can be retrieved via the Yahoo Finance API.

Data

The market data for this study was sourced from Yahoo Finance using Python, which served as the primary programming language throughout the analysis. Monthly closing prices were extracted for five firms: Apple Inc. (AAPL), Berkshire Hathaway Inc. (BRK-B), McDonald's Corporation (MCD), Toyota Motor Corporation (TM), and ExxonMobil (XOM). The sample period spans from March 2018 to March 2025. Monthly returns were calculated by applying the percent change function to the closing prices, resulting in a total of 85 observations. For the purposes of this research, an equally weighted portfolio comprising these five stocks is assumed. Descriptive statistics for the market returns are presented in Table 1. The data exhibits no extreme outliers; the highest monthly return is slightly above 19%, while the lowest is just below -21%. Figures 1 through 5 illustrate the return distributions for each stock, all of which appear approximately normally distributed.

[Table 1]

[Figure 1]

[Figure 2]

[Figure 3]

[Figure 4]

[Figure 5]

The primary sentiment data for this study was extracted and transformed from multiple sources. Initially, macro-level sentiment was intended to be gathered from social media platforms. However, due to API access constraints and web scraping limitations, the strategy was adjusted. Ultimately, the macro sentiment data was sourced from Federal Open Market Committee (FOMC) meeting transcripts spanning 2018 to 2025. These transcripts are publicly available via two separate URLs: fomc_historical_year and fomccalendars. Accessing both sites is necessary, as the transcripts from 2018–2019 are housed separately from those covering 2020–2025.

After retrieving the full-text meeting minutes, FinBERT, a financial domain-specific NLP model, was applied to extract sentiment scores from each transcript. Given that there are typically eight FOMC meetings per year, sentiment scores were aggregated and forward filled to produce monthly values. As a result, each month was assigned a positive, negative, and neutral sentiment score based on the FinBERT analysis.

Descriptive statistics for the sentiment data are reported in Table 2. Notably, the distribution of positive sentiment is centered around 1, while both negative and neutral sentiment scores cluster around 0. Figure 6 shows the distribution of positive sentiment, which is concentrated around 0.99 with a few outliers. Figure 7 presents the distribution of negative sentiment, also tightly clustered near 0. Figure 8 illustrates the distribution of neutral sentiment, which exhibits slightly greater dispersion but remains centered around 0.

[Table 2]

[Figure 6]

[Figure 7]

[Figure 8]

The second source of sentiment data was obtained from GDELT Project, a global open-access platform that continuously monitors broadcast, print, and online news from nearly every country in over 100 languages. GDELT provides rich datasets on a wide range of global themes, events, entities, and narratives. For this study, I utilized global macroeconomic sentiment derived from GDELT's archives.

Initially, I extracted all articles containing specific economic phrases that I identified as representative of macroeconomic themes. However, during the web scraping process, I encountered several months with little to no data due to the narrow scope of selected phrases. To address this, I revised the filter criteria by broadening the list of keywords, thereby allowing more articles to qualify for sentiment extraction.

Daily tone scores were extracted from each article and then averaged to compute monthly sentiment values. These monthly averages are presented as the avg_tone variable in Table 2, representing global macro sentiment. Figure 9 displays a histogram of avg_tone over the sample period, revealing a distribution that is generally symmetric but exhibits slight right skewness.

[Figure 9]

The remainder of the dataset was extracted to serve as features for the gradient boosted trees model. To enhance model interpretability, I incorporated key macroeconomic indicators from the Federal Reserve Economic Data (FRED) database: the Federal Funds Rate (fed_funds_rate), 10-Year Treasury Rate (treasury_10y_rate), Consumer Price Index (CPI), M2 Money Supply (m2_money_supply), Unemployment Rate (unemployment_rate), Real Gross Domestic Product (real_gdp), and Consumer Sentiment (consumer_sentiment). These variables collectively offer a snapshot of the broader macroeconomic environment, which can influence investor sentiment and market behavior.

By including these features, the model gains additional explanatory power, enabling us to assess which macro indicators are most influential in predicting market returns. Table 3 presents the descriptive statistics for the macroeconomic data. Notably, m2_money_supply and real_gdp stand out as extreme outliers relative to the scale of the other variables. Figures 10 through 16 display histograms of these features over the sample period. Of particular interest is Figure 14, which shows that unemployment_rate is highly right-skewed.

[Table 3]

[Figure 10]

[Figure 11]

[Figure 12]

[Figure 13]

[Figure 14]

[Figure 15]

[Figure 16]

After gathering the three types of data, some of the data within the macro data frame was back and forward filled to account for the difference in periods of aggregation. The data is ready for modeling.

Methodology

The methodology of this study draws inspiration from Kaffel et al. (2024), but introduces modifications to better suit the scope of this analysis. Rather than following their three-phase structure, I adopt a four-phase framework and omit the *Data Extraction* and *NLP Scoring* components, as these processes have already been addressed. The revised framework consists of: (1) Econometric Prediction, (2) Machine Learning Prediction, (3) Trading Simulation, and (4) Performance Analysis. This structure establishes the foundation for evaluating whether sentiment and macroeconomic indicators can be leveraged to forecast returns and construct profitable trading strategies.

ARIMA models are powerful statistical tools for predictive modeling with time series data. Researchers such as Rubio et al (2023) Implement ARIMA models to forecast volatility and stock prices. An ARIMA model consists of three components: Autoregressive (AR), moving average (MA) and non-stationary differences. The ARIMA model is as follows:

$$y_t = c + \sum_{k=1}^{p} \phi_k(y_{t-k}) + \sum_{k=1}^{q} \theta_k(\varepsilon_{t-k}) + \varepsilon_t$$

In the equation above, y_t represents the value of times series at time t. The next variable is constant c serves as the intercept. The autoregressive components use past values of the variable to predict its current value and are represented by $\sum_{k=1}^{p} \phi_k(y_{t-k})$. The moving average parts of the model use past forecast errors to improve current predictions and are represented by $\sum_{k=1}^{q} \theta_k(\varepsilon_{t-k})$. The unpredictable, white noise is denoted as ε_t.

To identify the best-fitting ARIMA model, I introduce the Akaike Information Criterion (AIC), a widely used metric in model selection. AIC balances model fit with complexity by penalizing excessive parameters, thereby helping prevent overfitting. Among competing models, the one with the lowest AIC is considered the best. The AIC is calculated using the formula below, where k represents the number of estimated parameters in the model and L is the maximized value of the likelihood function:

$$AIC = 2k - 2\ln(L)$$

The AIC found the best fitting model to be ARIMA(3,0,1). Furthermore, the best model includes three autoregressive lags, zero differencing terms, and one lagged forecasting error. The equation below represents our final baseline ARIMA model:

$$y_t = c + \phi_1(y_{t-1}) + \phi_2(y_{t-2}) + \phi_3(y_{t-3}) + \theta_1(\varepsilon_{t-1}) + \varepsilon_t$$

Our baseline ARIMA model does not include sentiment data. To incorporate sentiment into our econometric model, expansion is needed. The SARIMAX model extends a traditional ARIMA by allowing for seasonal components and accepting exogenous variables. To test whether sentiment features could improve forecasting monthly returns, I added my sentiment data as exogenous variables. The SARIMAX model is as follows.

$$y_t = c + \phi_1(y_{t-1}) + \phi_2(y_{t-2}) + \phi_3(y_{t-3}) + \theta_1(\varepsilon_{t-1}) + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \varepsilon_t$$

The new component in this equation are the exogenous variables and their coefficients denoted by $\beta_k x_{kt}$.

A powerful XGBoost model is also implemented to enhance predictive power. The premise of gradient boosting is to add new models sequentially, each one attempting to correct the errors made by its predecessor. This machine learning approach captures complex, non-linear interactions and accommodates a wide array of input features without degrading model performance. As an ensemble of decision trees, XGBoost iteratively optimizes to minimize a specified loss function. Conceptually, the model learns a function that maps macroeconomic and sentiment-based inputs to expected returns. The model can be written as:

$$\hat{y}_t = \sum_{m=1}^{M} f_m(X_t)\,, f_m \in \mathcal{F}$$

$\hat{y}_t$ represents the predicted return at time t. The sentiment and macro features are X_t. The individual regression trees are denoted by f_m. Lastly, $\mathcal{F}$ represents all possible trees.

The final component before performance analysis is a simulated portfolio. The objective of this simulation is to evaluate whether a sentiment-based trading strategy can outperform a market benchmark. Monthly closing prices for SPY, an ETF that tracks the S&P 500, were retrieved from Yahoo Finance, and returns were computed using percentage change. The trading strategy was straightforward: invest equally in all selected stocks in any month where the best-performing model forecasted positive returns. This simulation assumes perfect execution and does not account for transaction costs, slippage, or the fact that sentiment indicators are derived from historical macroeconomic and market data.

Results

In this section, I present the results from the modelization described in the methodology. Table 4 displays the output from the ARIMA(3,0,1) model. Both the first-order autoregressive term (ar.L1) and the first-order moving average term (ma.L1) are statistically significant at the 95% confidence level. The ar.L1 coefficient is −0.2036, suggesting that a 1% increase in returns in month t−1t - 1t−1 is associated with a 0.2036% decrease in returns in month t, holding other factors constant. Conversely, the ma.L1 coefficient of 0.9739 indicates that a 1% positive forecast error in the previous period leads to a 0.97% increase in the current period's return, demonstrating how unexpected shocks can substantially influence return dynamics. The remaining autoregressive and moving average lags were not statistically significant.

[Table 4]

[Figure 17]

Figure 17 displays the forecast made by the baseline ARIMA (3,0,1) 6 months in the future with a relatively opaque 95% confidence interval band. The fitted line is visually near identical to the historical line.

The results of the SARIMAX model, which incorporates sentiment variables as exogenous regressors, are shown in Table 5. Among the sentiment inputs, the coefficients for avg_tone*,* positive*,* and negative sentiment were statistically insignificant, suggesting limited standalone predictive value. However, the *neutral* sentiment variable exhibited a large and statistically significant negative coefficient of −1041.94, indicating a strong negative relationship with returns. This implies that an increase in neutral sentiment, as captured by FinBERT, is associated with a sharp decline in expected returns. Interestingly, the autoregressive and moving average terms that were significant in the baseline ARIMA model remained significant, reinforcing the robustness of those temporal dependencies. Figure 18 visualizes the SARIMAX model's six-month forecast, with the predictive range shaded in red to represent the 95% confidence interval.

[Table 5]

[Figure 18]

The final model implemented in this study was the XGBoost algorithm. Figure 19 displays a scatter plot of the actual versus predicted values, indicating that the model performed well, most predictions lie close to the 45-degree line, suggesting minimal deviation from actual returns. Additionally, Figure 20 presents a time-series comparison of actual and predicted returns. The predicted return series closely tracks the actual return series, reinforcing the model's ability to capture the underlying patterns in the data with reasonable accuracy.

[Figure 19]

[Figure 20]

Extreme gradient boosted trees typically offer improved predictive accuracy at the expense of interpretability. Nonetheless, many features were included in the model to enhance explanatory power. Figure 21 presents the top 10 most important features as determined by the model. Interestingly, consumer sentiment emerged as the most influential predictor, originally sourced from macroeconomic data rather than the sentiment dataset. Other key features include the unemployment rate, ranked second, and positive sentiment, which ranked third. Notably, while positive sentiment was statistically insignificant in the SARIMAX model, it became a critical feature under XGBoost. Neutral sentiment ranked seventh, trailing both positive and negative sentiment in importance, suggesting that nonlinear models may uncover interaction effects or hidden patterns not captured by linear frameworks.

[Figure 21]

Table 6 presents the comparative performance of the three models evaluated in this study. The selection of evaluation metrics was informed by Atsalakis (2009), who surveyed various stock market forecasting techniques and emphasized that soft computing methods are widely accepted for analyzing and assessing market behavior.

The first metric, Root Mean Squared Error (RMSE), penalizes large prediction errors more heavily and is commonly minimized in forecasting models. The ARIMA model produced an RMSE of 0.02366, with the SARIMAX model yielding a slightly higher value of 0.02386. The XGBoost model outperformed both, achieving a lower RMSE of 0.0188, indicating superior forecasting accuracy.

Mean Absolute Error (MAE), which captures the average magnitude of forecast errors regardless of direction, followed a similar pattern. The ARIMA model posted an MAE of 0.0176, marginally outperforming SARIMAX (0.0179). XGBoost again outperformed both, with the lowest MAE of 0.0141.

R-squared (R^2) measures the proportion of variance in returns explained by the model. While ARIMA and SARIMAX yielded comparable R^2 values of 0.4145 and 0.4045 respectively, XGBoost demonstrated a significantly better fit, with an R^2 of 0.6205.

Lastly, Directional Accuracy, defined as the percentage of time the model correctly predicted the direction of return, highlighted a clear strength of the machine learning approach. XGBoost achieved a directional accuracy of 87.05%, far surpassing SARIMAX (74.12%) and ARIMA (72.94%).

[Table 6]

Given that XGBoost outperformed both the ARIMA and SARIMAX models across all evaluation metrics, the simulated trading strategy is based on the signal generated by the XGBoost forecasts. Figure 22 displays cumulative returns beginning in March 2018, assuming the investor follows a straightforward rule: take a long position when the model predicts a positive return for the upcoming month. The strategy assumes an equal-weighted allocation across the five stocks analyzed in this study.

Also included in Figure 22 is the performance of the SPY benchmark over the same period. As shown in Table 7**,** the sentiment-based strategy consistently outperformed the benchmark across every performance metric. The Sharpe Ratio for the strategy was 0.8154**,** compared to 0.2323 for the benchmark, indicating higher risk-adjusted returns. The maximum drawdown for the strategy was only -0.0423, significantly lower than the benchmark's -0.2393, suggesting greater downside protection.

The hit rate**,** or percentage of time the strategy correctly predicted the direction of returns, was 72.94%**,** outperforming the benchmark's 66.66%**.** One of the most striking results came from the Profit Factor, which compares the total profit of winning trades to the total loss of losing trades. The sentiment strategy achieved a Profit Factor of 12.751, vastly exceeding the benchmark's 1.7603**,** highlighting the robustness and potential profitability of a sentiment-driven trading approach.

[Figure 22]

[Table 7]

Conclusion

Predicting stock prices remains a compelling challenge that continues to attract interest from both investors and academics. This study investigated the nuanced relationship between macroeconomic sentiment and stock market returns by employing multiple modeling techniques and evaluating their comparative performance. The findings offer a valuable contribution to the intersection of econometric modeling and machine learning in financial forecasting. More specifically, this work serves as a useful resource for researchers examining how macroeconomic sentiment influences broader market dynamics. A fully documented Python script has been developed to ensure the study's reproducibility. The remainder of this section discusses the key insights derived from the analysis, as well as the broader implications of the results, limitations of the approach, and directions for future research.

FOMC-derived sentiment demonstrates predictive power, particularly when the sentiment is classified as *neutral*, which exhibits a negative relationship with returns. Through FinBERT analysis, sentiment scores were extracted from FOMC transcripts; however, neutrality emerged as the only statistically significant sentiment indicator. This suggests that neutral language induces investor indecision, often prompting selling behavior. Notably, the SARIMAX model, despite incorporating sentiment, only marginally outperformed the baseline ARIMA in directional accuracy. This implies that sentiment-based signals are more effective for directionally driven strategies rather than for precise return forecasting. In practice, the optimal application of FOMC sentiment involves responding to neutrality shocks by either shorting the market or, for more risk-averse investors, using neutrality as a signal to avoid initiating long positions, ideally in conjunction with other trading signals.

Sentiment derived from the GDELT Project did not exhibit any statistically significant impact on returns across the models. This outcome is likely attributed to the high degree of noise and variability present in the extracted sentiment data. For future research, not only would a larger dataset be beneficial, but it is also essential to ensure that the extracted data is more targeted and directly relevant to the research question.

Machine learning models were benchmarked against hybrid time series econometric models to forecast stock returns using macroeconomically derived sentiment. Strategic integration of sentiment indicators and macroeconomic variables enhanced the predictive power of each model. XGBoost leveraged a wide array of these features, many of which were also tested in econometric models, to deliver superior performance. Across all evaluation metrics, including error minimization, goodness of fit, and directional accuracy, XGBoost outperformed the alternatives, accurately predicting the direction of returns nearly 87% of the time.

A significant challenge in this study was the limited dataset, only 85 monthly observations spanning seven years. This short time horizon introduces a potential bias that may have affected the robustness of the model forecasts. Additionally, the trading simulation incorporated sentiment data through 2025, which introduces look-ahead bias by allowing the model to act on future information, an unrealistic assumption in real-world trading environments. To improve the validity of future models, acquiring historical sentiment data from premium sources may be necessary, as freely available archives of internet-based sentiment are limited. Furthermore, incorporating real-time sentiment would enable a more accurate assessment of the long-term viability of sentiment-based trading strategies.

In conclusion, this study highlights the potential of integrating sentiment analysis with both econometric and machine learning models to improve stock return forecasting. The findings suggest that while sentiment alone may not consistently yield statistically significant results, its strategic incorporation, particularly through models like XGBoost, can enhance predictive accuracy and directional insight. By addressing the challenges and limitations encountered in this research, including data constraints and forward-looking bias, future studies can develop more robust and reliable models. Overall, sentiment analysis remains a promising tool in the arsenal of quantitative finance, with the capacity to refine market predictions and trading strategies.

References

Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2019). *Taking the Fed at its word: A new approach to estimating central bank objectives using text analysis*. *American Economic Journal: Macroeconomics, 11*(4), 1–32. https://doi.org/10.1093/restud/rdab094

Kim, Y., & Kim, K. (2014). *Investor sentiment from internet message postings and predictability of stock returns*. *Journal of Economic Behavior & Organization, 107*, 708–729. https://doi.org/10.1016/j.jebo.2014.04.015

Rubio, L., Romero-Meza, R., & Pino, G. (2023). *Forecasting volatility by using wavelet transform, ARIMA and GARCH models: The case of Tesla stock*. *Physica A: Statistical Mechanics and its Applications, 616*, 128683. 10.1007/s40822-023-00243-x

Atsalakis, G. S., & Valavanis, K. P. (2009). *Surveying stock market forecasting techniques –*

Part II: Soft computing methods. *Expert Systems with Applications, 36*(3), 5932–5941. https://doi.org/10.1016/j.eswa.2008.07.006

Fama, E. F., & French, K. R. (1988). *Permanent and temporary components of stock prices*. *Journal of Political Economy, 96*(2), 246–273. https://www.jstor.org/stable/1833108

Panpoonsup, S., Anantasunthorn, T., & Pichitlamken, J. (2022). *Evaluating the effectiveness of sentiment-based models for stock price prediction*. *Journal of Financial Data Science, 4*(3), 54–68. 10.2139/ssrn.4185507

Hentati Kaffel, Rania and Ben Yahia, Sarra and Garcia Sanchez, José Angel, Impact of Sentiment Analysis on Energy Sector Stock Prices: A Finbert Approach. Available at SSRN: https://ssrn.com/abstract=5018148 or http://dx.doi.org/10.2139/ssrn.5018148

Oh, C., & Sheng, O. R. L. (2011). *Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement*. *International Conference on Information Systems (ICIS)*, 1–17. paper

Supporting Tables

Table 1: Descriptive Statistics – Market Data

	Count	Mean	Std	Min	Max
avg_tone	85	2.2052705	0.05238143	2.09586402	2.31644307
positive	85	0.99999069	3.43E-05	0.99973589	0.99999917
neutral	85	3.59E-06	1.06E-05	4.35E-08	8.22E-05
negative	85	5.71E-06	2.37E-05	5.57E-07	0.00018181

Table 2: Descriptive Statistics –Sentiment Data

	Count	Mean	Std	Min	Max
avg_tone	85	2.2052705	0.05238143	2.09586402	2.31644307
positive	85	0.99999069	3.43E-05	0.99973589	0.99999917
neutral	85	3.59E-06	1.06E-05	4.35E-08	8.22E-05
negative	85	5.71E-06	2.37E-05	5.57E-07	0.00018181

Table 3: Descriptive Statistics –Macro Data

	Count	Mean	Std	Min	Max
Fed_Funds_Rate	85	2.386	1.99955364	0.05	5.33
Treasury_10Y_Rate	85	2.68933981	1.23694178	0.62363636	4.79809524
CPI	85	279.544	24.2604969	249.529	319.775
M2_Money_Supply	85	18789.4	2943.46978	13919.3	21749.6
Unemployment_Rate	85	4.61294118	2.04146237	3.4	14.8
Real_GDP	83	21628.2801	1183.06542	19056.617	23542.349
Consumer_Sentiment	85	78.1788235	14.6357365	50	101.4

Table 4: ARIMA (3,0,1) Summary Table

Var	Coefficient	Std. Error	z	P>|z|	95% CI Lower	95% CI Upper
const	0.0137	0.005	2.994	0.003	0.005	0.023
ar.L1	-0.2036	0.094	-2.175	0.03	-0.387	-0.02
ar.L2	-0.2263	0.14	-1.619	0.106	-0.5	0.048
ar.L3	0.0871	0.166	0.524	0.6	-0.238	0.412
ma.L1	0.9739	0.094	10.322	0	0.789	1.159
sigma2	0.0005	0.0001	6.599	0	0	0.001

Table 5: SARIMAX Summary Table

Variable	Coefficient	Std. Error	z	P>|z|	95% CI Lower	95% CI Upper
avg_tone	0.0066	0.073	0.09	0.928	-0.136	0.149
positive	0.0004	0.161	0.002	0.998	-0.316	0.317
neutral	-1041.9434	304.053	-3.427	0.001	-1637.875	-446.011
negative	528.6263	675.422	0.783	0.434	-795.177	1852.43
ar.L1	-0.2063	0.099	-2.083	0.037	-0.4	-0.012

ar.L2	**-0.2518**	**0.148**	**-1.697**	**0.09**	**-0.543**	**0.039**
ar.L3	**0.1015**	**0.171**	**0.593**	**0.553**	**-0.234**	**0.437**
ma.L1	**1.0216**	**0.126**	**8.128**	**0**	**0.775**	**1.268**
sigma2	**0.0005**	**0**	**3.888**	**0**	**0**	**0.001**

Table 6: Model Comparisons

Model	RMSE	MAE	R2	Directional_Accuracy
ARIMA	**0.0236644**	**0.01760658**	**0.41458436**	**72.94117647**
SARIMAX	**0.02386644**	**0.01794302**	**0.40454584**	**74.11764706**
XGBoost	**0.01880006**	**0.01418501**	**0.63051925**	**87.05882353**

Table 7: Sentiment Metrics vs Benchmark

Metrics	Sentiment Strategy	SPY Benchmark
Sharpe Ratio	**0.8154**	**0.2323**
Max Drawdown	**-0.0423**	**-0.2393**
Hit Rate (%)	**72.9412**	**66.6667**
Profit Factor	**12.751**	**1.7603**

Supporting Figures

Figure 1: Histogram of Apple returns

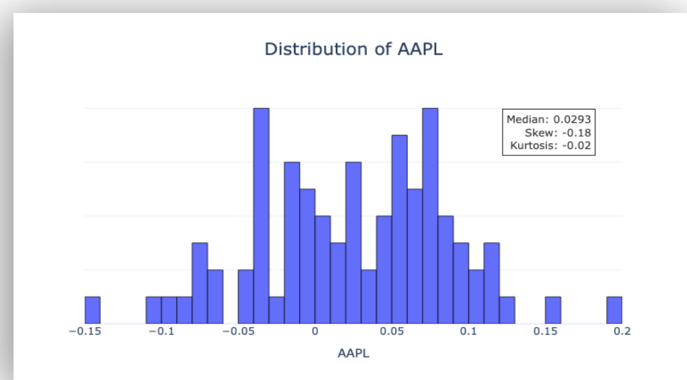


Figure 2: Histogram of Berkshire Hathaway returns

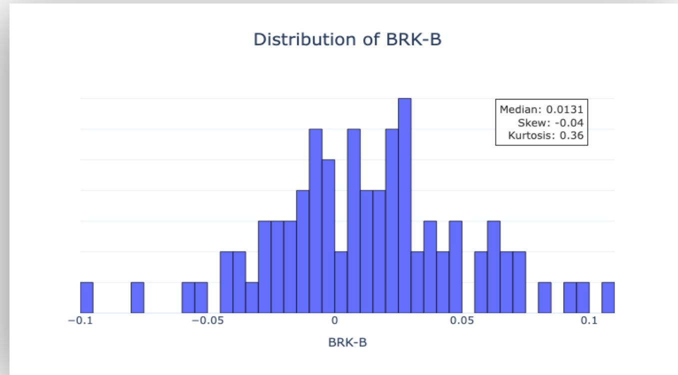


Figure 3: Histogram of McDonald's returns

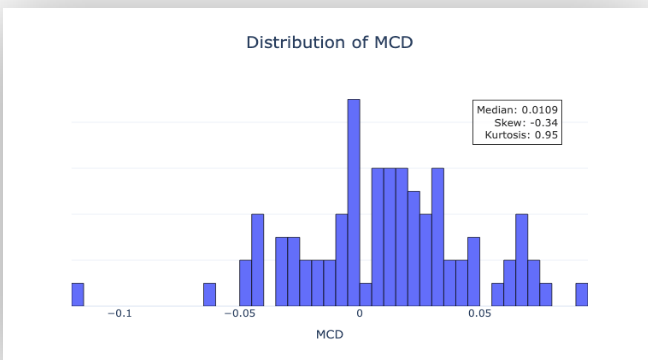


Figure 4: Histogram of Toyota returns

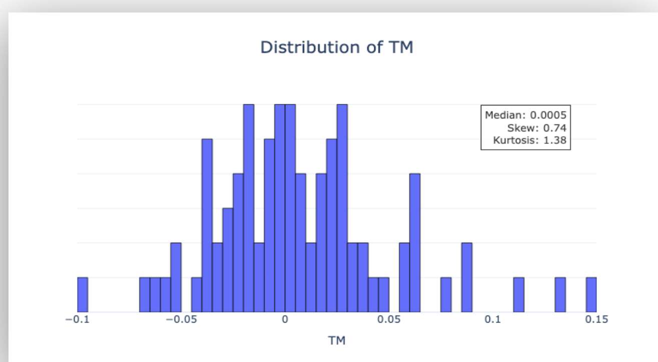


Figure 5: Histogram of Exxon returns

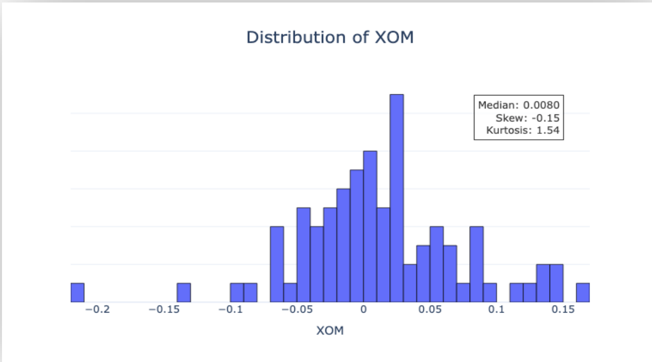


Figure 6: Histogram of Positive Sentiment

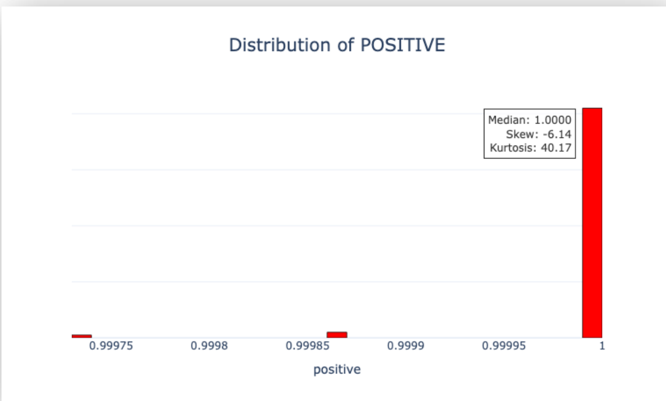


Figure 7: Histogram of Negative Sentiment

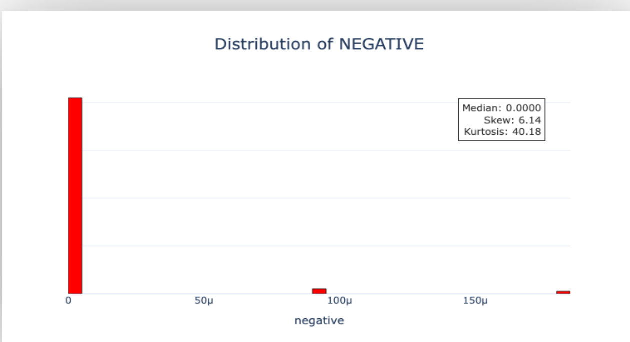


Figure 8: Histogram of Neutral Sentiment

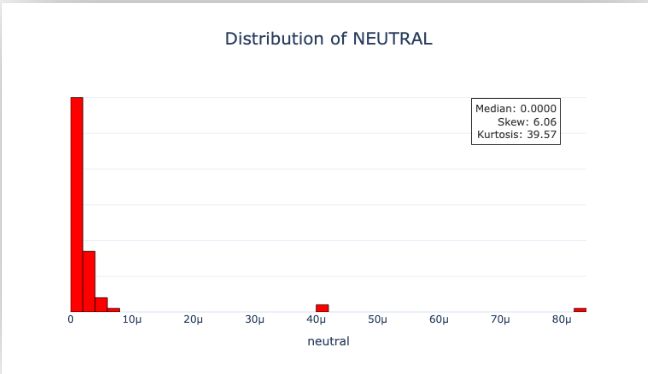


Figure 9: Histogram of Global Macro Sentiment (avg_tone)

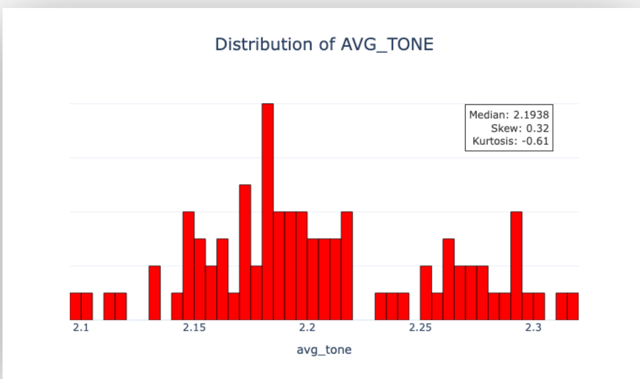


Figure 10: Histogram of Federal Funds Rate

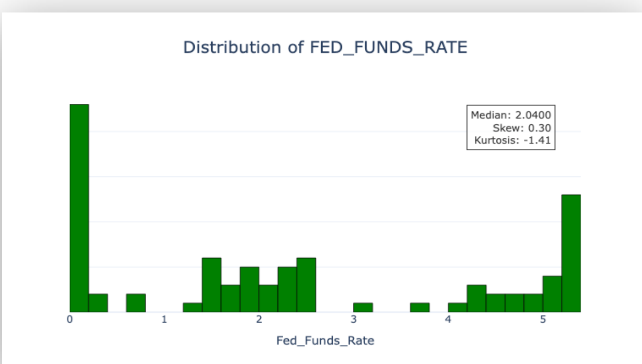


Figure 11: Histogram of 10-Year treasury Rate

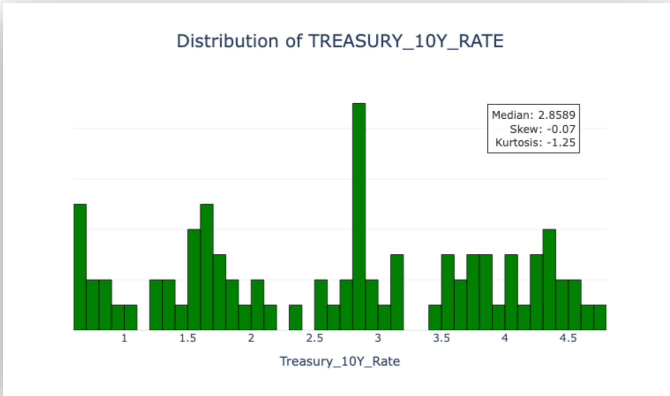


Figure 12: Histogram of Consumer Price Index

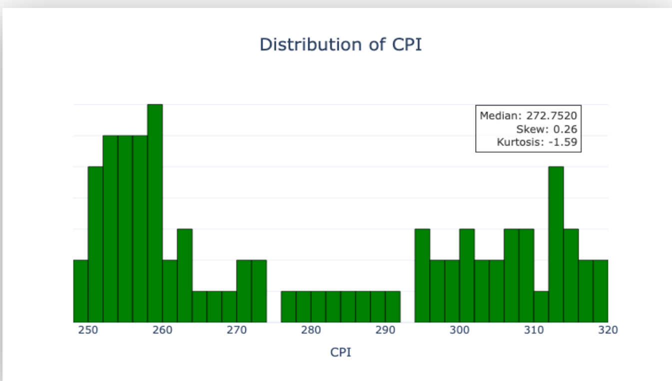


Figure 13: Histogram of M2 Money Supply

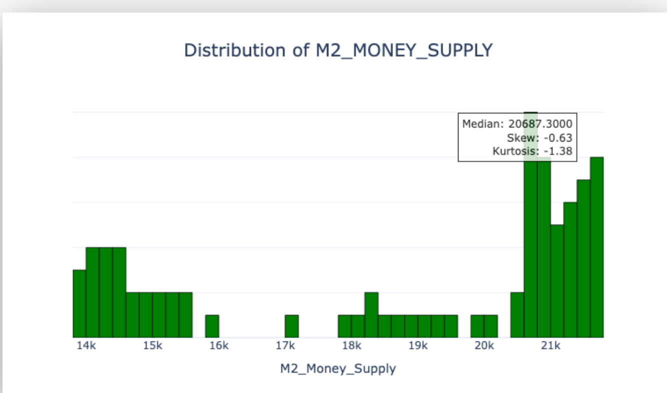


Figure 14: Histogram of Unemployment Rate

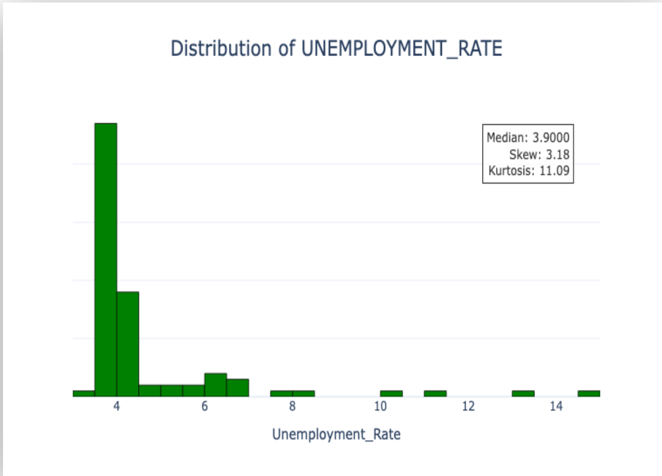


Figure 15: Histogram of Real Gross Domestic Product

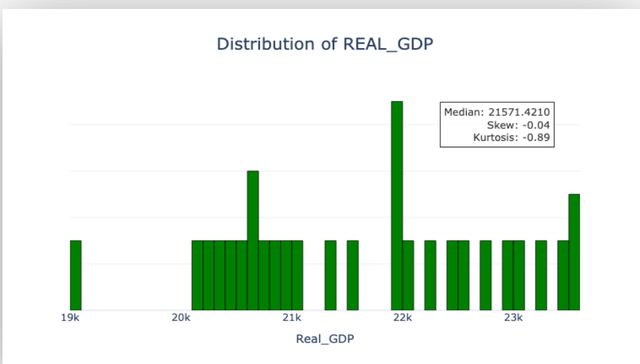


Figure 16: Histogram of Consumer Sentiment

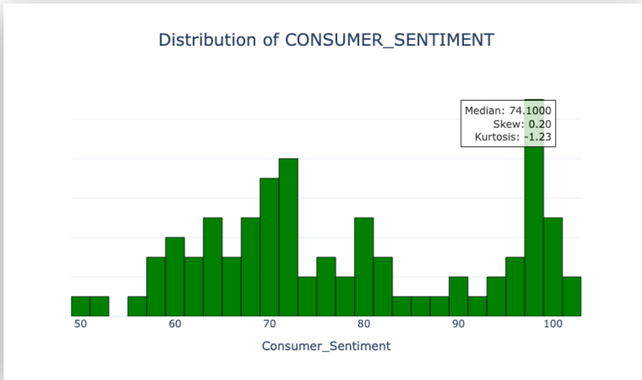


Figure 17: ARIMA (3,0,1) 6 Month Forecasts

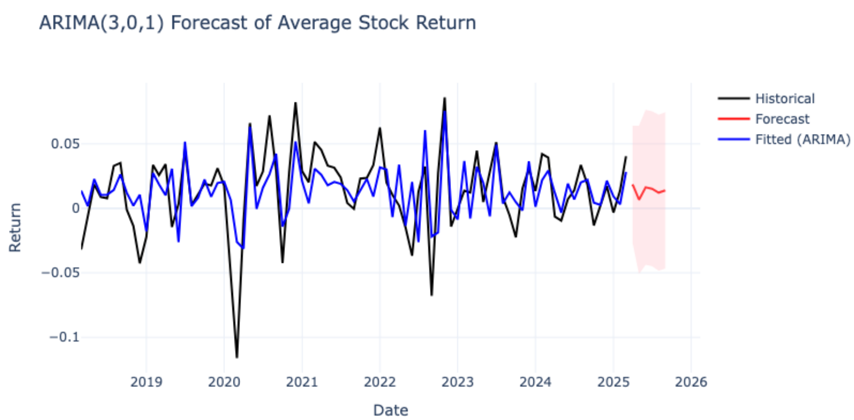


Figure 18: SARIMAX 6 Month Forecasts

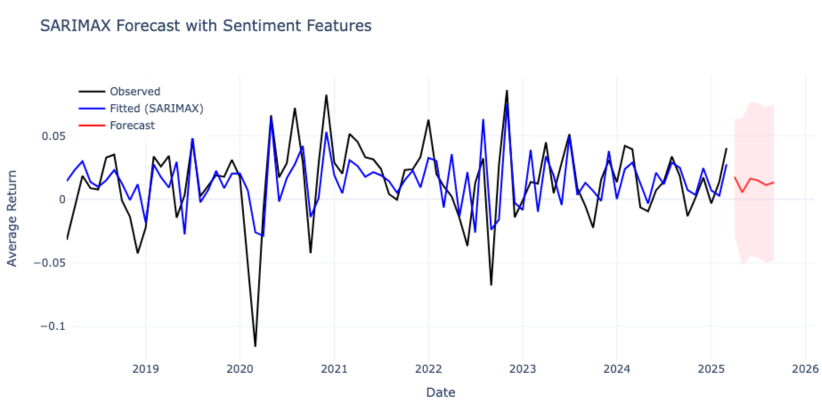


Figure 19: Scatter Plot of XGBoost predictions

Figure 20: XGBoost Predictions vs Actual Returns

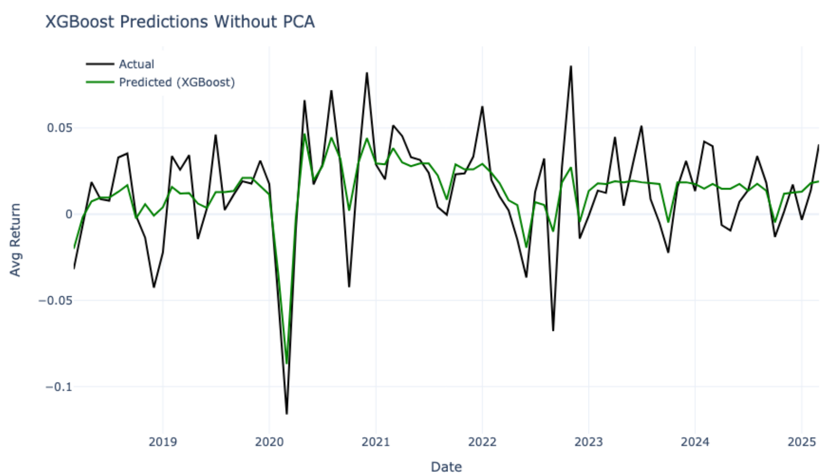


Figure 21: Top 10 Feature Importance

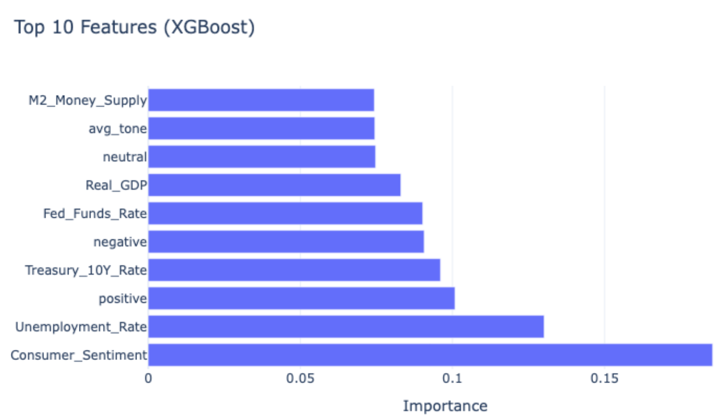
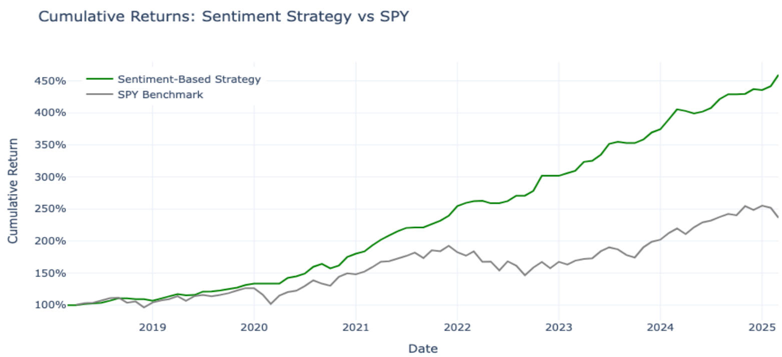


Figure 22: Cumulative Returns



CODE: