

Predicting the impact on Stock Market with Current Trade Tariff

Announcements

Saakshi Dedhia, Saumya Jhaveri, Xinqi Li, and Siyu Chen

Professor Matthew Delventhal - ECON 570

May 14th, 2025

Abstract

This study investigates how financial news related to the Trump tariffs, from March 31 to April 19, 2025, influences the direction of short-term stock market returns (positive or negative) at both the company and ETF levels. Using a dataset of 706 financial news articles collected from reliable sources such as Yahoo Finance and CNN, we apply FinBERT to extract 768-dimensional semantic embeddings that capture latent textual signals. At the company level, we align these embeddings with minute level stock return data for 14 firms across various sectors. We use LASSO as a filtering method to select the most predictive features, which are then used as input variables for XGBoost classification to predict stock return directions at the company level. Model performance is evaluated using accuracy, AUC, and ROC analysis. In parallel, we analyze three ETF markets—DIA, SPY, and XLI—using minute-level data to compare model performance across modeling approaches, including XGBoost and Random Forest. Models are assessed based on out-of-sample accuracy, AUC, F1 score, and other metrics. Results show that specific FinBERT dimensions—particularly *finbert_59*—consistently contribute to return prediction across companies. However, the accuracy of predicting company-level stock return directions is lower than that at the ETF level. This may be due to the greater complexity and heterogeneity of individual companies, which are influenced by a wider range of firm-specific factors. In contrast, ETF-level performance is notably stronger, with both XGBoost and Random Forest models demonstrating high predictive power in forecasting ETF return directions.

Keywords: Financial Text Analysis; Stock Return Direction Prediction; FinBERT; LASSO Feature Selection; XGBoost Classification; Random Forest

1. Introduction

The trade policies implemented during President Donald Trump’s administration, particularly the announcement and imposition of tariffs on major trading partners such as China, the European Union, and others, triggered significant reactions across global financial markets. These policy moves reshaped investor sentiment, introduced volatility, and led to notable shifts in market performance. Headlines often emerged within minutes, followed by rapid swings in asset prices, reflecting heightened market sensitivity to political and economic uncertainty. Motivated by these observations, we seek to quantify market-level responses to policy shocks through a data-driven framework. The broader relevance of this inquiry is clear: in a globally interconnected economy, a single policy announcement by a major power can propagate ripples across international stock markets within hours. Understanding these dynamics is critical not only for academic inquiry but also for informing investors’ portfolio allocation and hedging strategies, as well as offering policymakers insight into how financial markets digest uncertainty and shifts in global power structures. Accordingly, our project integrates event-driven data, text-based sentiment analysis, and financial modeling to capture and measure these immediate market responses.

Building on this motivation, we investigate the relationship between financial news sentiment and short-term stock return directions at two levels: individual company stocks and three major exchange-traded funds (ETFs)—DIA, SPY, and XLI. To extract meaningful textual features, we leverage FinBERT, a financial-domain-specific adaptation of BERT, trained on financial corpora (Araci, 2019). Each news article is transformed into a 768-dimensional embedding vector, allowing a rich capture of nuanced semantic information beyond simple sentiment classification. We employ the full FinBERT embeddings to preserve the latent

structure inherent in financial language. Using LASSO, we filter the top five most influential latent features associated with stock return movements, which are then used as input variables for XGBoost to predict stock return directions. Moreover, by extending the analysis to major ETFs, we evaluate whether sentiment patterns influence broader market indices, thereby providing macro-level validation of the impact of financial news on ETF return directions. Finally, we examine the accuracy of predictions for stock return directions at both the company and ETF levels in order to evaluate the relative effectiveness of our models across different market scopes.

2. Literature Review

Recent academic work has increasingly incorporated sentiment analysis into financial modeling frameworks. For example, Wenjun Gu et al. (2024) introduced a hybrid FinBERT-LSTM model that combined financial news sentiment with historical stock price trends to forecast NASDAQ-100 movements. By integrating more than 800,000 news articles from Benzinga with LSTM sequences, their model achieved an accuracy rate as high as 95.5%, outperforming traditional LSTM and deep neural network baselines. Their findings illustrate a broader shift in financial econometrics, where traditional time series models are increasingly augmented with sentiment-driven features, particularly during periods of policy-driven market volatility.

A growing body of research has further refined how financial sentiment is measured. Araci (2019) introduced FinBERT, a financial-domain-specific adaptation of BERT, pretrained on financial corpora to capture the subtle nuances of financial language. FinBERT has since become a standard tool for financial text classification, often outperforming general-purpose language models. Building on this foundation, Zou and Herremans (2022) extended FinBERT's application by using full sentence embeddings to predict extreme Bitcoin price movements, demonstrating that dense textual representations could capture latent predictive signals. Similarly, Shobayo et al. (2024) compared FinBERT and GPT-4 embeddings for stock trend prediction, showing that embeddings could enhance forecasting beyond traditional sentiment scoring.

Inspired by these observations, our study aims to analyze the impact of financial news on stock return directions. We integrate FinBERT embeddings, LASSO-based feature selection, and

apply XGBoost and Random Forest models to predict return directions at both the company and ETF levels. By comparing model performance across these two levels, our approach provides a comprehensive view of how sentiment influences different segments of the market.

3. Data Collection

3.1 Stock Market Data

This study investigates the relationship between financial news sentiment and short-term stock return directions at two levels: company-level equities and broad-market exchange-traded funds (ETFs), specifically DIA, SPY, and XLI. The sample period spans from March 31 to April 19, 2025, a window chosen to capture market behavior during a phase of heightened economic uncertainty and significant trade policy developments.

This timeframe includes key political and economic events such as the introduction of a 10% baseline tariff in early April, and the subsequent implementation of a 145% tariff targeting Chinese imports. The introduction of a baseline tariff, and the implementation of targeted import tariffs.

The selected period allows for the observation of immediate and cumulative market reactions, aiming to reduce the influence of confounding factors such as corporate earnings announcements and broader macroeconomic cycles, thereby providing a relatively controlled setting for high-frequency financial analysis.

3.1.1 ETF Minute-Level Data

Minute-by-minute trading data for the SPDR Dow Jones Industrial Average ETF (DIA), SPDR S&P 500 ETF (SPY), and the Industrial Select Sector SPDR Fund (XLI) were obtained from the LSEG workspace. The Industrial Select Sector SPDR Fund (XLI) is an exchange-traded fund that tracks the performance of the industrial sector of the S&P 500 Index. It includes large-cap U.S. companies involved in manufacturing, construction, transportation, aerospace, defense, and related industries. Key constituents of XLI typically include firms like Boeing,

Caterpillar, Honeywell, and Union Pacific. These ETFs were chosen to capture both broad market sentiment (DIA, SPY) and sector-specific dynamics (XLI).

The XLI (Industrial Select Sector SPDR Fund) specifically tracks the performance of the industrial sector within the S&P 500 index, including companies involved in aerospace and defense, machinery, transportation, construction, and manufacturing. It aggregates stocks from some of the largest and most influential industrial firms, such as Boeing, Caterpillar, 3M, Union Pacific, and Honeywell, making it particularly sensitive to changes in trade policies, economic growth, and infrastructure spending.

The datasets span approximately three weeks of continuous minute-level observations, providing detailed information on price movements, trading volume, and transaction counts.

During the sample period, trading activity across all three ETFs demonstrated concentrated liquidity within a relatively narrow price range, accompanied by active intraday fluctuations.

3.1.2 Company-Level Minute-Level Data

Company-specific stock price data were obtained via the Alpha Vantage API, focusing on major publicly traded firms with significant exposure to international trade and global supply chains.

To capture detailed intraday price movements while managing noise, the stock prices were aggregated into minute-level intervals. This approach aligns closely with the timing of key financial news releases, providing a more precise measure of market sensitivity. The selected companies cover a diverse range of sectors, including technology, industrials, and consumer discretionary, as listed below:

Table 3.1.2. Selected Firms for Company-Level Analysis

Ticker	Company	Sector
AAPL	Apple Inc.	Technology
TSLA	Tesla Inc.	Consumer Discretionary
INTC	Intel Corporation	Technology
AMD	Advanced Micro Devices Inc.	Technology
NVDA	NVIDIA Corporation	Technology
F	Ford Motor Company	Consumer Discretionary
GM	General Motors Company	Consumer Discretionary
WHR	Whirlpool Corporation	Consumer Discretionary
NKE	Nike Inc.	Consumer Discretionary
CAT	Caterpillar Inc.	Industrials
BA	The Boeing Company	Industrials
BABA	Alibaba Group Holding Limited	Technology (International)
JD	JD.com Inc.	Technology (International)
PDD	PDD Holdings Inc.	Technology (International)

Note. The table presents the selected firms for company-level stock analysis, categorized by their corresponding sectors. International companies are separately identified.

The selection of firms at the individual level allows for a more detailed analysis of market dynamics, capturing both broad market movements and sector-specific responses. This approach provides the flexibility to isolate the impact of macroeconomic events, such as trade policy shifts or earnings announcements, on specific industries like technology and consumer discretionary, which are particularly sensitive to economic conditions. By including both domestic and international firms, this dataset also reflects the global nature of modern financial markets, enhancing the relevance and robustness of the analysis.

3.2 Text Data

For text data collection, we accessed eventregistry.org for news filtering where we filtered based on the ‘Trump speech’ term. We downloaded the sheet and extracted the full body of articles with a python script. Initially, we extracted a total of 908 articles. After the text cleaning process, we retained 706 useful articles. Some of these articles contained duplicated content in the text, but differed in source titles, unique identifiers (URIs), and recorded sentiment scores. The articles cover the period from March 31st to April 19th. Sources include major financial media outlets such as Yahoo Finance, IndustryWeek, and others, focusing primarily on critical financial news. The timestamps associated with each article are recorded with precision down to the second.

Following the data cleaning, we constructed a dataset consisting of 706 articles with duplicated text but differing sentiment scores and source information. Our objective is to extract rich textual information by applying the FinBERT model, using its pre-trained package to transform the articles into 768-dimensional embeddings. This allows us to capture deep semantic features.

3.3 Time Alignment

To support both company-level and broader market analyses, we aggregated news articles and stock returns into minute-level blocks, enabling us to accommodate the uneven timing of news arrivals and trading activity. For individual companies, stock return data—originally available at an hourly frequency via the Alpha Vantage API—was aligned with the minute-level news blocks to ensure temporal consistency. In contrast, for ETFs, which exhibit higher liquidity and more continuous trading behavior, we retrieved minute-level stock price data to achieve finer alignment with article timestamps. Market-wide news articles were similarly timestamped and

synchronized with the minute-by-minute ETF return series, allowing us to more accurately capture sentiment-driven market reactions in real time.

4. Data Methodology

This project aims to predict short-term stock market reactions using financial news articles collected from reputable sources such as Yahoo Finance, CNN, BNN, TradingEconomics, Coindesk, Cryptonews, BBC, and others. By aligning news texts with corresponding stock market data, we employ FinBERT—a domain-specific language model trained on financial text—to extract latent linguistic features from the articles. These features are then used to model and forecast stock return movements, offering insights into how market behavior is shaped by the informational and emotional content embedded in financial news. Focusing on the event-driven context of the Trump tariff policy shocks, we compile high-frequency financial news alongside matching stock return data. Through the integration of textual feature extraction and predictive modeling, our methodology investigates how financial markets—at both the company and ETF levels—respond to sentiment-laden textual signals in terms of return direction.

4.1 Feature Extraction

For the 706 unique financial news articles, we apply FinBERT to extract semantic embeddings. Each article is processed through the FinBERT model, producing a 768-dimensional vector that captures the latent semantic features of the text. These embeddings are compiled into a feature matrix, with each row representing an article and each column corresponding to a specific embedding dimension.

Table 4.1. Summary of Selected News Articles and their Sentiment Distribution

Total_Articles <int>	Unique_Time_Blocks <int>	Avg_Word_Count <dbl>	Positive_Perc <dbl>	Neutral_Perc <dbl>	Negative_Perc <dbl>
706	614	893.8	42.4	23.1	34.6

Table 4.1 presents descriptive statistics for the company-level news dataset used in this study. A total of 706 financial articles were collected and aligned with 1-minute stock return intervals, spanning 614 unique time blocks. The average article length is 893.8 words, reflecting the detailed nature of the financial reporting. In terms of sentiment composition, 42.4% of the articles are classified as positive, 34.6% as negative, and 23.1% as neutral.

4.2 Modelling Framework

a. Company-Level Stock Modeling:

To assess the predictive value of textual features on company-level stock return directions, we use LASSO to select the top five embedding features, which serve as input variables in an XGBoost model. This enables us to evaluate the predictive power of the most influential features.

First, we apply LASSO regression to the FinBERT embedding feature matrix. LASSO performs variable selection. The nonzero coefficients identified in the LASSO models reveal which latent semantic features are most closely associated with stock return variability.

LASSO Equation for Variable Selection:

$$\hat{\beta} = \arg \min_{\beta} \left\{ \sum_{i=1}^n \left(y_i - \sum_{j=1}^{768} X_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^{768} |\beta_j| \right\}$$

y_i is the observed minute level stock rate of return, X_{ij} is the j th FinBERT feature for observation i , β is the regression coefficient for feature j , and λ is the regularization parameter. By identifying a sparse set of influential FinBERT embedding dimensions, LASSO enables a more interpretable and robust linkage between financial news text and stock market movements.

The rate of return for a given time block (4-hour) is calculated as:

$$\text{Return}_t = \frac{P_{\text{close},t} - P_{\text{open},t}}{P_{\text{open},t}}$$

Among the selected features, we further identify the top feature for each company, defined as the embedding dimension with the largest absolute coefficient. This represents the strongest predictor of stock return variation within the FinBERT embedding space for that company. To evaluate predictive performance, we calculate several standard accuracy metrics based on out-of-sample predictions using λ_{min} . These include: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2). These metrics are reported in a summary table for all 14 companies to facilitate cross-comparison of model fit and performance across firms. In addition, to further explore the informational content of the selected features, we identify the top five features for each company based on the magnitude of their LASSO coefficients (as summarized later in the Results section).

Second, to classify the direction of stock price movements (positive or negative returns), we train XGBoost classifiers using the FinBERT embeddings as input features. Classification performance is evaluated based on accuracy and the area under the receiver operating characteristic curve (AUC), providing a comprehensive assessment of model discrimination ability.

To frame our binary classification task, we define the target variable as:

$$\text{Target} = \begin{cases} 1 & \text{if } r_c > 0 \\ 0 & \text{otherwise} \end{cases}$$

Where r_c represents the stock return for a given company. This transformation allows us to model the direction of market movement, predicting whether a return will be positive or not. The XGBoost model outputs a predicted probability \hat{y}_i for each time block i , indicating the estimated likelihood that the return is positive. A hard label is then assigned by thresholding this probability:

$$\hat{y}_{\text{label}} \leftarrow I[\hat{y} > 0.5]$$

where I is the indicator function. The XGBoost model builds its prediction iteratively using gradient boosting. At each boosting round t , a new decision tree $f_t(x_i)$ is trained to minimize the residual errors from the previous model stage, and the prediction is updated according to:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta f_t(x_i)$$

Here, η denotes the learning rate (set to 0.1 in our case), which controls the contribution of each new tree. This iterative optimization continues until convergence or early stopping, resulting in a robust ensemble model that balances prediction accuracy and generalization.

Through this dual modeling framework, we aim to assess not only which latent features are important predictors but also how well rich text-based features can classify return directions compared to traditional sentiment measures.

b. ETF Markets Stock Modeling:

To evaluate the predictive value of sentiment dynamics on the direction of daily stock market returns, we employ a supervised machine learning framework focused on classification. Specifically, we train Extreme Gradient Boosting (XGBoost) classifiers using sentiment-informed feature vectors derived from FinBERT-based textual analysis and technical market indicators. The modeling pipeline is structured as follows:

Sentiment Feature Construction:

Sentiment data is sourced from financial news articles processed via FinBERT, a transformer-based language model fine-tuned on financial corpora. For each trading day, we compute aggregated sentiment indicators, including:

1. **Mean Sentiment Score:** Captures the average sentiment polarity across articles.
2. **Sentiment Volatility (Standard Deviation):** Reflects the heterogeneity of sentiment.
3. **Article Volume:** Total count of sentiment-tagged news articles per day.

These features serve as proxies for market tone and information intensity, capturing both directional sentiment and its dispersion.

Market Feature Engineering

Historical price data (Open, High, Low, Close, Volume) is transformed into derived indicators such as:

1. **Daily Return:** Computed as the percentage change in closing price.
2. **Direction Label:** A binary outcome variable indicating positive (1) or negative (0) return movement.
3. **Volatility Measures:** Range-based volatility and rolling return deviations.
4. **Volume Changes:** Captures liquidity shifts.

The resulting dataset combines market-based and sentiment-based predictors for each day.

Temporal Data Integration

To align the sentiment and market datasets, we apply a soft temporal join using nearest-neighbor matching within a ± 1 -day window. This approach accommodates timing mismatches between news publication and trading activity, ensuring realistic data availability assumptions and avoiding forward-looking bias.

Model Training and Classification

We train an XGBoost classifier on the integrated feature set to predict the binary return direction (up/down) for each day. XGBoost is a gradient-boosted tree ensemble model known for its scalability, ability to handle non-linear feature interactions, and resistance to overfitting through regularization.

Key aspects of the training process include:

1. **Train-Test Split:** Time-aware partitioning is used to avoid lookahead bias.
2. **Feature Scaling:** Predictors are standardized using StandardScaler or RobustScaler.
3. **Hyperparameter Tuning:** XGBoost's learning rate, depth, and regularization parameters are optimized using grid search or default tuning heuristics.

Evaluation Metrics

Model performance is assessed using multiple standard metrics, including:

1. **Accuracy:** Proportion of correctly predicted labels.
2. **Precision & Recall:** Class-wise sensitivity and specificity.
3. **F1 Score:** Harmonic mean of precision and recall.
4. **AUC-ROC:** Area under the receiver operating characteristic curve, measuring model discrimination capability.

These metrics are reported for both training and test sets to evaluate generalizability.

Feature Importance Analysis

Following model training, XGBoost's built-in feature importance rankings are used to identify which predictors—whether sentiment-based or market-based—most significantly impact model decisions. This analysis provides insights into the relative informational value of textual sentiment compared to traditional price-based indicators.

5. Analysis and Empirical Results

5.1 Big Picture of All Text data

We collected text data related to financial news about the Trump tariffs from March 31 to April 19, 2025. As an overview of the text corpus (see Appendix, Graph 1), the most frequently occurring words include 'tariffs', 'Democrats', 'party', 'trade', 'Republicans', 'leaders', 'US', etc. In terms of article frequency over time (see Appendix, Graph 2), we observe a significant spike around the tariff announcement, with over 300 articles published in response to Trump's speech. The highest volume of news coverage occurred on April 1 and April 2. These patterns highlight the critical role of political framing and media intensity in shaping investor sentiment, providing a rationale for using time-sensitive sentiment features to predict short-term return directions.

5.2 Company-level Analysis and Results

To examine how stock return directions are influenced by financial news, we use LASSO to select the top features from the 768 FinBERT embeddings. These selected embeddings represent the most relevant textual signals affecting company stock returns. We then input these features into an XGBoost classification model to evaluate their predictive power. The following sections present the results of our variable selection and XGBoost modeling.

5.2.1 LASSO Variable Selection Results

The following table (Table 5.2.1) presents the outcomes of this feature selection process.

Company	Lambda_Min	Top_Feature	Top_Coefficient	Total_Rows	MSE	RMSE	MAE	R2
AAPL	0.000902	finbert_315	0.002580239	350	0.00013412	0.011581	0.006539	0.272928
TSLA	0.001823	finbert_59	-0.005156399	350	0.000243725	0.015612	0.009808	0.204064
AMD	0.002242	finbert_59	-0.003528313	350	0.000276135	0.016617	0.00895	0.117797
NVDA	0.002095	finbert_59	-0.00296422	350	0.000212892	0.014591	0.009313	0.146665
F	0.001381	finbert_59	-0.003051132	350	9.85854E-05	0.009929	0.006262	0.111774
GM	0.001254	finbert_156	0.001936534	350	0.000106614	0.010325	0.006157	0.110567
CAT	0.00136	finbert_59	-0.002903062	350	0.000107795	0.010382	0.005572	0.107548
BA	0.001541	finbert_59	-0.004955799	350	0.000197209	0.014043	0.007901	0.158636
BABA	0.001631	finbert_634	0.003090211	350	0.000180959	0.013452	0.008018	0.103988
PDD	0.001927	finbert_29_lag1	0.001936769	350	0.000160973	0.012688	0.00709	0.05723

Table 5.2.1. Summary of LASSO Filtering Results for Company-Level Stock Return Prediction

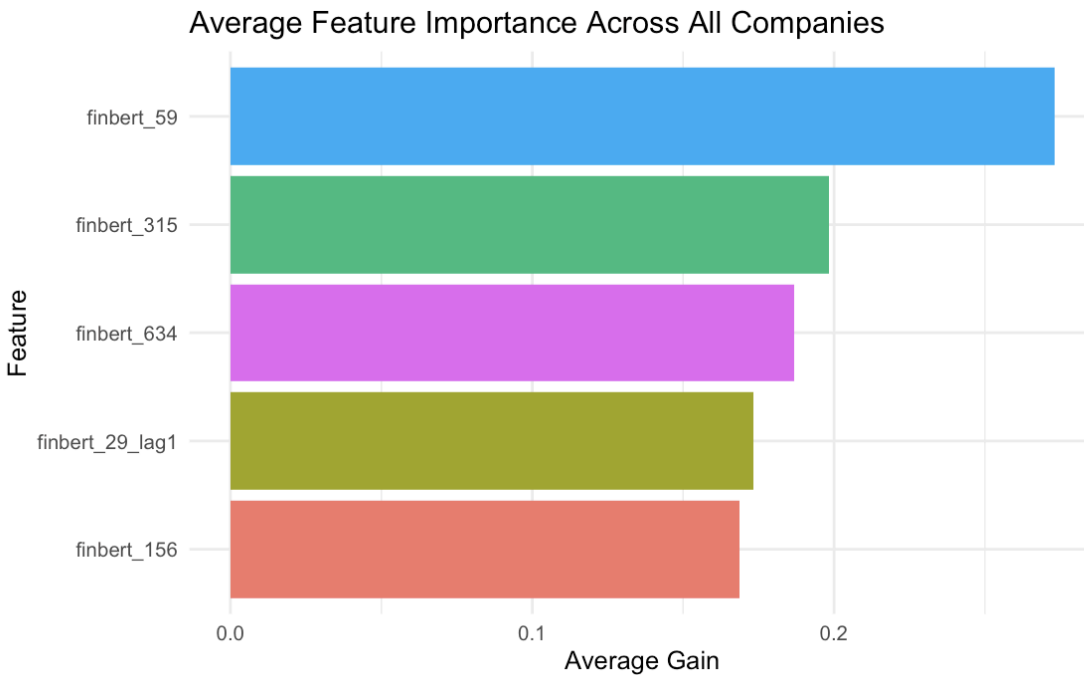
As shown above, for each of the companies analyzed, the optimal regularization parameter (λ) was determined through 5-fold cross-validation, resulting in company-specific λ values. The top feature for each company—defined as the FinBERT embedding dimension with the largest absolute LASSO coefficient—varies across firms. Notably, finbert_59 emerges as the most frequently selected top feature, appearing in models for TSLA, AMD, NVDA, F, CAT, and BA. This recurrence suggests that finbert_59 captures a latent semantic signal that is broadly relevant across firms, although the direction and magnitude of its influence differ by company. Other top features, such as finbert_156 (GM), finbert_634 (BABA), and lagged dimensions like finbert_29_lag1 (PDD), highlight the heterogeneous ways in which firms respond to sentiment-laden financial information.

The accuracy metrics indicate moderate predictive performance. RMSE values range from approximately 0.0099 to 0.0166, and R^2 values span from 0.057 (PDD) to 0.273 (AAPL),

suggesting that the FinBERT-based models explain a limited but meaningful portion of return variability. While no company exceeds an R^2 of 0.3, firms such as AAPL and TSLA show relatively stronger explanatory power. In contrast, companies like PDD and BABA exhibit weaker predictability, indicating that sentiment alone may not fully capture price movements for these stocks. Overall, the findings underscore the potential of using textual embeddings to forecast short-term market reactions, while also revealing the importance of tailoring models to firm-specific dynamics.

The graph of Optimal Lambda (min) by Company is provided in the Appendix.

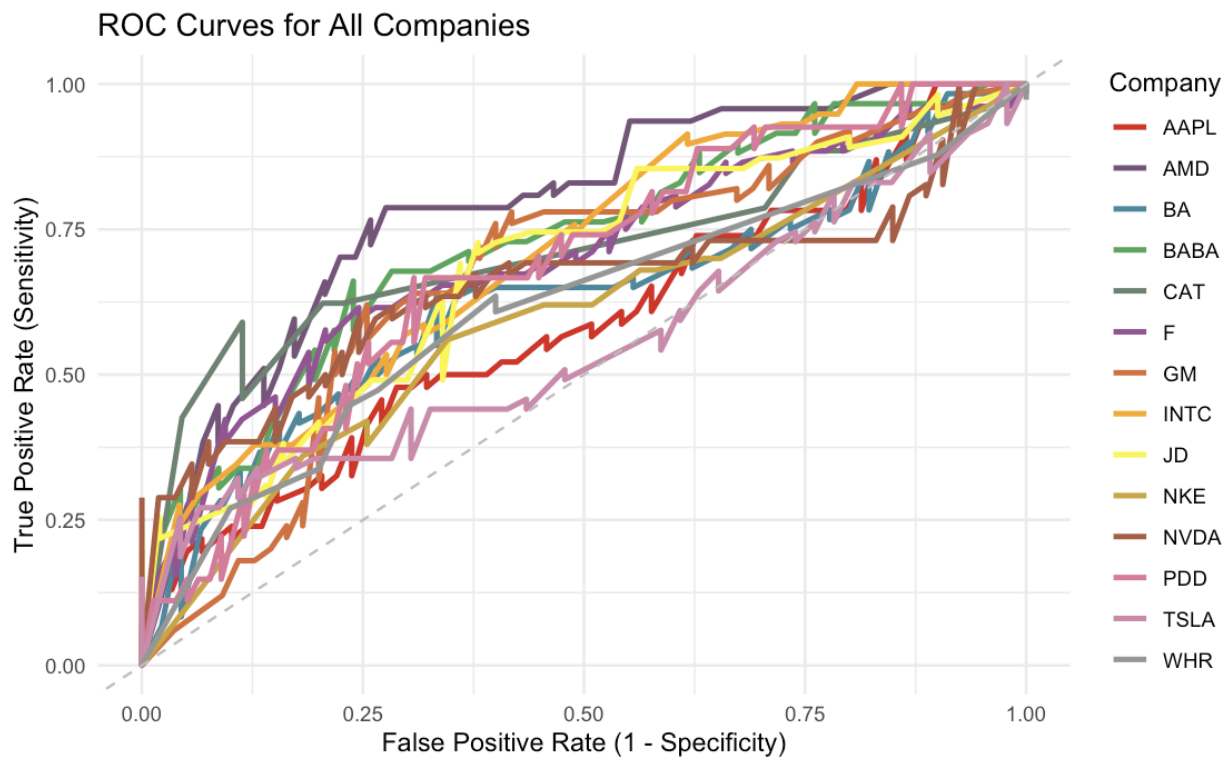
5.2.2 XGBoost Results and Accuracy Test



Graph 5.2.2.a. Average XGBoost Feature Importance Across All Companies (Top 5 FinBERT Dimensions)

Graph 5.2.2.a illustrates the average feature importance of the top five FinBERT dimensions—identified through LASSO selection—across all companies, based on XGBoost

classification models trained to predict stock return directions. The y-axis lists the most influential features, while the x-axis shows their average gain, representing each feature's contribution to improving model accuracy during boosting iterations. Among these, `finbert_59` stands out as the most dominant predictor, demonstrating a consistently high average gain across firms. This suggests that `finbert_59` captures a robust latent semantic signal that generalizes well across different market contexts. Other important features include `finbert_315`, `finbert_634`, `finbert_29_lag1`, and `finbert_156`, indicating that both contemporary and lagged textual signals play substantial roles in driving market reactions. Additionally, an individual-level comparison of different companies' feature importance is put in Appendix .



Graph 5.2.2.b. ROC Curves for XGBoost Classifiers Across 14 Companies

Graph 5.2.2.b presents the Receiver Operating Characteristic (ROC) curves for the XGBoost classification models trained on each of the companies in the dataset. Each curve

illustrates the performance of a binary classifier tasked with predicting the direction of stock returns (positive vs. negative), based on FinBERT-derived semantic features and minute-level time block context.

The ROC curve plots the True Positive Rate (Sensitivity) against the False Positive Rate ($1 - \text{Specificity}$) across varying classification thresholds. The diagonal dashed line represents a random guess baseline ($\text{AUC} = 0.5$). Curves that rise above this line indicate better-than-random performance.

In this analysis, nearly all company models show ROC curves above the diagonal, confirming that the FinBERT-augmented features provide meaningful predictive signals. Companies such as AMD, JD, and TSLA exhibit notably steeper curves, especially in the early portion of the x-axis, reflecting stronger sensitivity at lower false positive rates. On the other hand, flatter curves for firms like PDD and WHR suggest weaker classification power, likely due to noisier sentiment signals or less pronounced price reactions to news.

These results highlight the firm-specific variability in how sentiment-driven text features impact classification performance and suggest that while FinBERT embeddings are generally helpful, their predictive strength varies by firm.

Additional supporting materials are provided in the Appendix, including: (1) a graph of the *Top 5 Feature Importances by Company* based on XGBoost gain scores, (2) a *Confusion Matrix* summarizing the XGBoost model's performance across all companies, (3) a table presenting *Company-Level XGBoost Classification Performance* in terms of accuracy and AUC scores by firm, and (4) a graph comparing *Accuracy and AUC Scores* of XGBoost classifiers across companies. These materials provide further insight into model performance and feature-level interpretability.

5.3 ETF-Level Analysis and Results

To evaluate how effectively our model predicts daily stock return direction based on financial and sentiment-derived features, we implemented a single but robust classification approach using XGBoost (Extreme Gradient Boosting). This model was applied to market data enriched with sentiment indicators for a selected index or asset. XGBoost was chosen for its ability to handle structured data, capture non-linear feature interactions, and mitigate overfitting through regularization.

A comprehensive set of evaluation metrics was employed to assess model performance, including:

- Accuracy: Overall proportion of correct directional predictions.
- Precision and Recall: To evaluate the correctness and completeness of positive class predictions.
- F1 Score: To balance precision and recall in a unified metric.
- AUC (Area Under the ROC Curve): To quantify the model's ability to distinguish between upward and downward market movements.
- RMSE (Root Mean Squared Error) and R^2 (Coefficient of Determination): To evaluate the calibration of probability estimates and the strength of the fit.

These metrics enable a multidimensional assessment of both classification quality and probabilistic reliability.

Rather than employing explicit feature selection techniques like LASSO, we trained the model on a curated set of 17 engineered features. These features include sentiment mean and

volatility, article count, return momentum, volume shifts, and volatility-based price indicators. Despite the inclusion of all features, the XGBoost model exhibited strong generalization performance, indicating that the engineered feature space was well-constructed and free from significant redundancy or noise.

5.3.1 XLI Results

For the industrial ETF XLI, XGBoost achieved an accuracy of 89.77%, with a precision of 91.41%, recall of 85.86%, and an F1 score of 88.55%. Notably, the model produced an AUC of 0.9692, confirming its excellent ability to discriminate between positive and negative return days. With an R^2 value of 0.5883 and RMSE of 0.3198, the model also demonstrated strong generalization and probability calibration on unseen data.

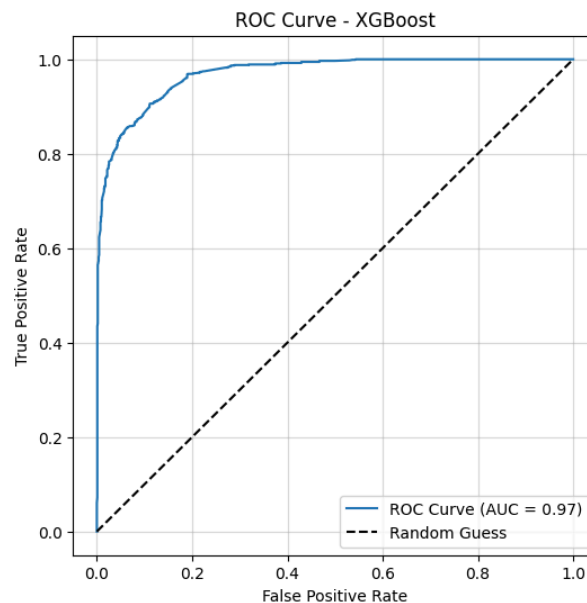


Image 5.3.1.a: ROC Curves for XGBoost for XLI

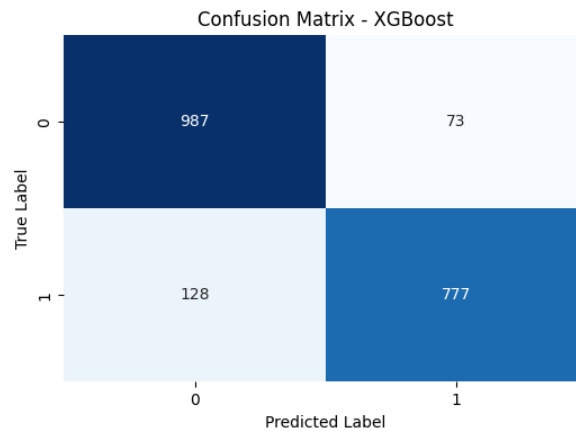


Image 5.3.1.b: Confusion Matrix for XGBoost for XLI

5.3.2 SPY Results

The SPY ETF delivered the best overall results among the three. XGBoost achieved an accuracy of 91.04%, a precision of 90.24%, recall of 90.67%, and a balanced F1 score of 90.46%. The AUC again reached 0.9692, indicating consistent and robust predictive discrimination. The R^2 of 0.6402 and RMSE of 0.2993 further reinforce the model's strong explanatory power and precision in estimating return direction.

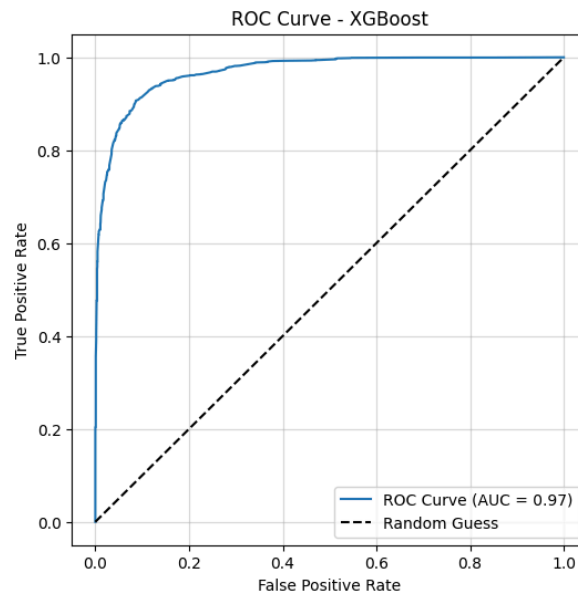


Image 5.3.2.a: ROC Curves for XGBoost for SPY

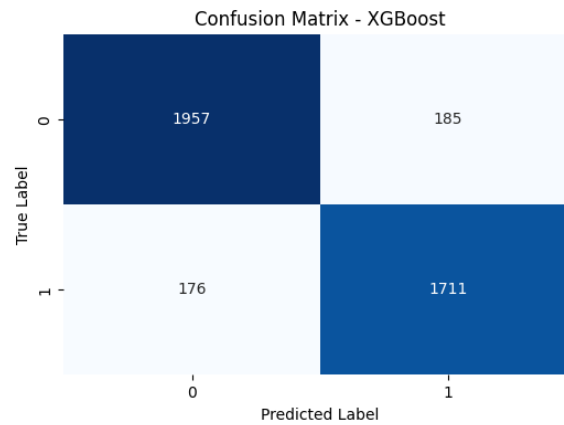


Image 5.3.2.b: Confusion Matrix for XGBoost for SPY

5.3.3 DIA Results

Performance on the DIA ETF was somewhat lower compared to XLI and SPY but still showed respectable results. XGBoost achieved an accuracy of 79.68%, with a precision of 83.18%, recall of 71.29%, and an F1 score of 76.77%. The AUC value of 0.8854 reflects solid classification capacity, although the R^2 score of 0.1843 suggests that DIA's price movements were less predictable—possibly due to its smaller number of constituents and greater sensitivity to individual stock-level volatility. The RMSE for DIA was 0.4508, indicating slightly noisier predictions compared to the other two ETFs.

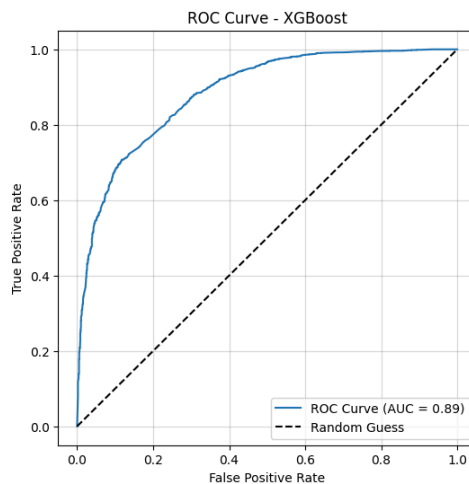
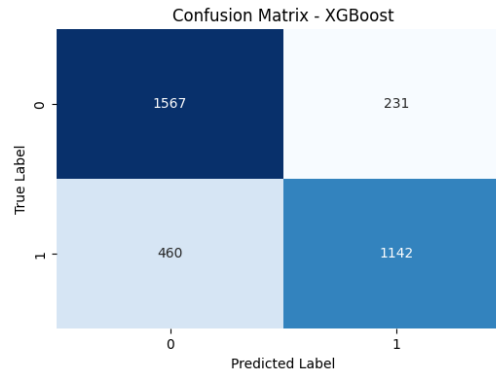


Image 5.3.3.a: ROC Curves for XGBoost for DIA**Image 5.3.3.b: Confusion Matrix for XGBoost for SPY**

Across all three ETFs, XGBoost proved to be a highly effective model for capturing the complex, nonlinear interactions between sentiment-derived and financial variables. Its consistent performance—especially in SPY and XLI—highlights the relevance of integrating textual sentiment analysis with technical market indicators for return direction forecasting. The confusion matrices further validate the model’s classification quality, showing a balanced distribution of false positives and false negatives across classes. Additionally, the high AUC scores across all datasets demonstrate the model’s robustness in distinguishing between positive and negative return classes.

These findings confirm that, for daily ETF prediction tasks, tree-based ensemble models such as XGBoost are well-suited, especially when paired with carefully engineered features combining structured financial data and unstructured textual sentiment. The results underscore the practical value of sentiment-aware machine learning in enhancing traditional predictive models for financial market behavior.

5.3.4 Summary

ETF	Model	Accuracy	Precision	Recall	F1 Score	AUC	RMSE	R ²
XLI	XGBoost	0.8977	0.9141	0.8586	0.8855	0.9692	0.3198	0.5883
SPY	XGBoost	0.9104	0.9024	0.9067	0.9046	0.9692	0.2993	0.6402
DIA	XGBoost	0.7968	0.8318	0.7129	0.7677	0.8854	0.4508	0.1843

Table 3: ETF-Level Model Performance Summary table

The superior performance of the XGBoost model can be attributed to its ability to flexibly capture complex, nonlinear interactions between engineered financial and sentiment-based features. XGBoost's ensemble tree-based structure excels in handling high-dimensional, structured data without requiring explicit assumptions about data distribution or temporal dependencies. This makes it particularly well-suited for financial prediction tasks, where market behavior is influenced by a combination of technical indicators, macroeconomic shocks, and sentiment-driven fluctuations. Given the daily frequency of the data and the inherently noisy and reactive nature of stock market movements, models that rely on sequential memory—such as recurrent neural networks—may not perform as effectively. In contrast, XGBoost directly leverages feature importance and robust decision boundaries, offering both interpretability and strong predictive performance in environments shaped by heterogeneous and often abrupt market signals.

Conclusion

This study investigates how financial news related to the Trump tariffs (March 31–April 19, 2025) influences stock return directions at both the company and ETF levels. Leveraging FinBERT embeddings to capture latent textual signals from financial articles, we utilize LASSO for feature selection and apply XGBoost and Random Forest models to classify returns as positive or negative. Rather than making explicit trading predictions, our primary goal is to evaluate and understand model accuracy across two distinct market scopes: individual companies and aggregated ETFs.

Results demonstrate that ETF-level predictions achieve higher accuracy compared to those at the company level. This likely reflects the broader and more uniform market sentiment captured by ETFs, whereas individual companies are influenced by diverse and complex firm-specific factors. In sum, this research highlights the potential of combining financial NLP with advanced machine learning techniques, emphasizing the need for flexible, enriched modeling frameworks to address complex market dynamics.

Despite these promising insights, several limitations exist, including the sparse distribution of textual data, coarse temporal granularity, potential survivorship bias in article selection, and exclusive reliance on FinBERT-derived features. Future research could mitigate these limitations by integrating additional data sources such as social media sentiment or macroeconomic indicators, employing finer temporal resolutions, and developing hybrid architectures that combine deep learning with tree-based modeling approaches.

References

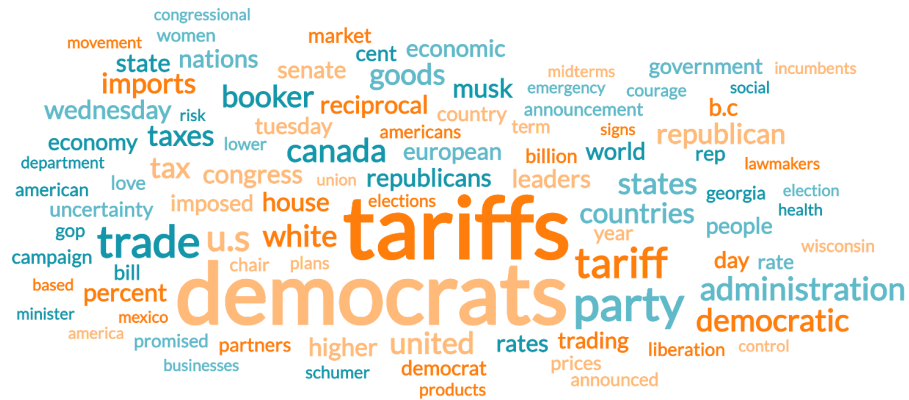
- Araci, D. (2019). FinBERT: Financial Sentiment Analysis with Pre-trained Language Models.
<https://arxiv.org/abs/1908.1006>
- Carhil. (n.d.). *Tariff's impacts on markets and portfolios* | BlackRock. BlackRock.
<https://www.blackrock.com/us/financial-professionals/insights/tariffs-and-investment-portfoliofolios>
- Cfa, C. L. S. (2025, April 9). How do tariffs impact the stock market? Kiplinger.com.
<https://www.kiplinger.com/investing/how-do-tariffs-impact-the-stock-market>
- Fajgelbaum, P., Khandelwal, A., & Maximilian Schwarz. (2021). *THE ECONOMIC IMPACTS OF THE US-CHINA TRADE WAR* (Gene Grossman, Jonathan Vogel, & Stephen Redding, Eds.; Working Paper 29315). NATIONAL BUREAU OF ECONOMIC RESEARCH. https://www.nber.org/system/files/working_papers/w29315/w29315.pdf
- Gu, W. J., Zhong, Y. H., Li, S. Z., Wei, C. S., Dong, L. T., Wang, Z. Y., & Yan, C. (2024). Predicting Stock Prices with FinBERT-LSTM: Integrating News Sentiment Analysis. *ArXiv*, 67–72. <https://doi.org/10.1145/3694860.3694870>
- How tariffs are forecast to affect US stocks. (2025, February 7). *Goldman Sachs*.
<https://www.goldmansachs.com/insights/articles/how-tariffs-are-forecast-to-affect-us-stocks>
- Kirsch, J. (n.d.). *How may tariffs affect the stock market and your investments?* Forbes.
<https://www.forbes.com/sites/investor-hub/article/how-tariffs-affect-stock-market-investments/>
- Morgan, J. (n.d.). *US tariffs: What's the impact?* | J.P. Morgan Research.
<https://www.jpmorgan.com/insights/global-research/current-events/us-tariffs>

Shobayo, O., et al. (2024). Innovative Sentiment Analysis and Prediction of Stock Price Using FinBERT, GPT-4, and Logistic Regression. *Big Data and Cognitive Computing*, 8(1), 143.

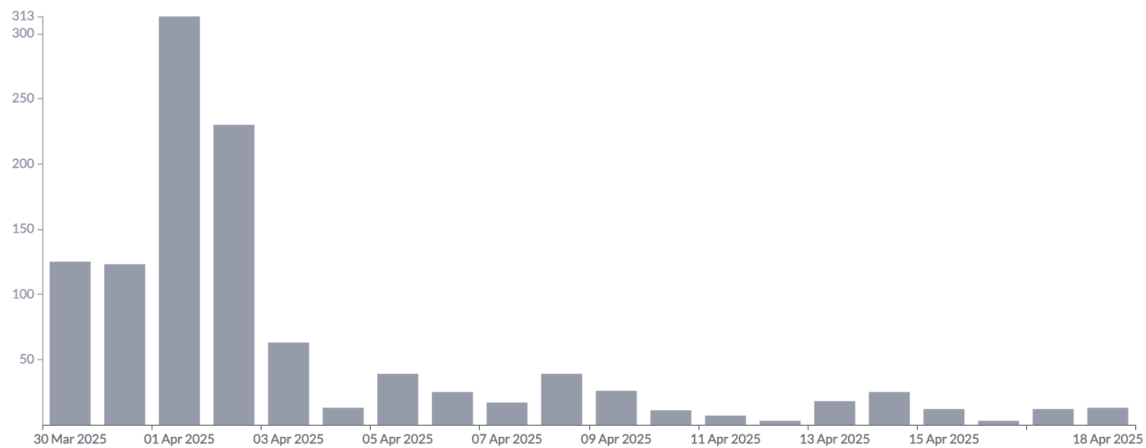
Warren, J. (2025, March 24). *Economist assesses tariff impacts on stock market, trade*. News. <https://news.ucr.edu/articles/2025/03/24/economist-assesses-tariff-impacts-stock-market-trade>

Zou, C., & Herremans, D. (2022). A Multimodal Model with Twitter FinBERT Embeddings for Extreme Price Movement Prediction of Bitcoin. <https://arxiv.org/abs/2206.00648>

Appendix

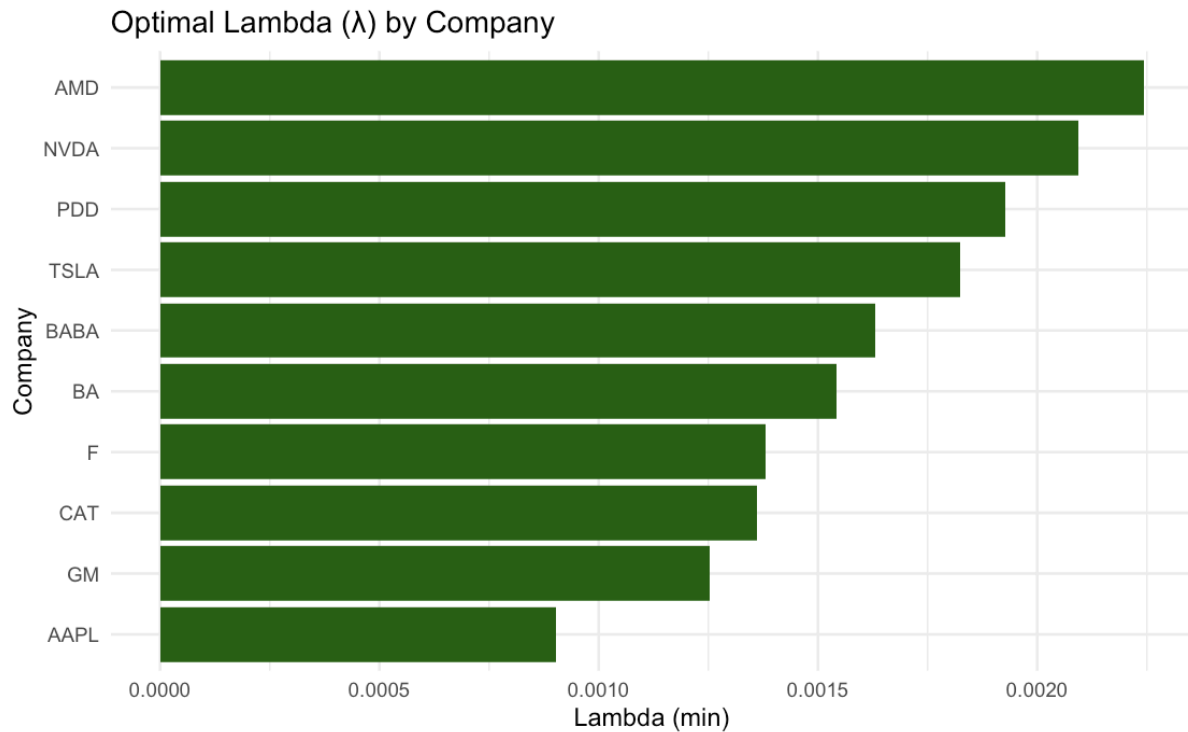


Graph 1: Word cloud for data from 3/31/2025-4/19/2025.



Graph 2: Article Timeline Frequency

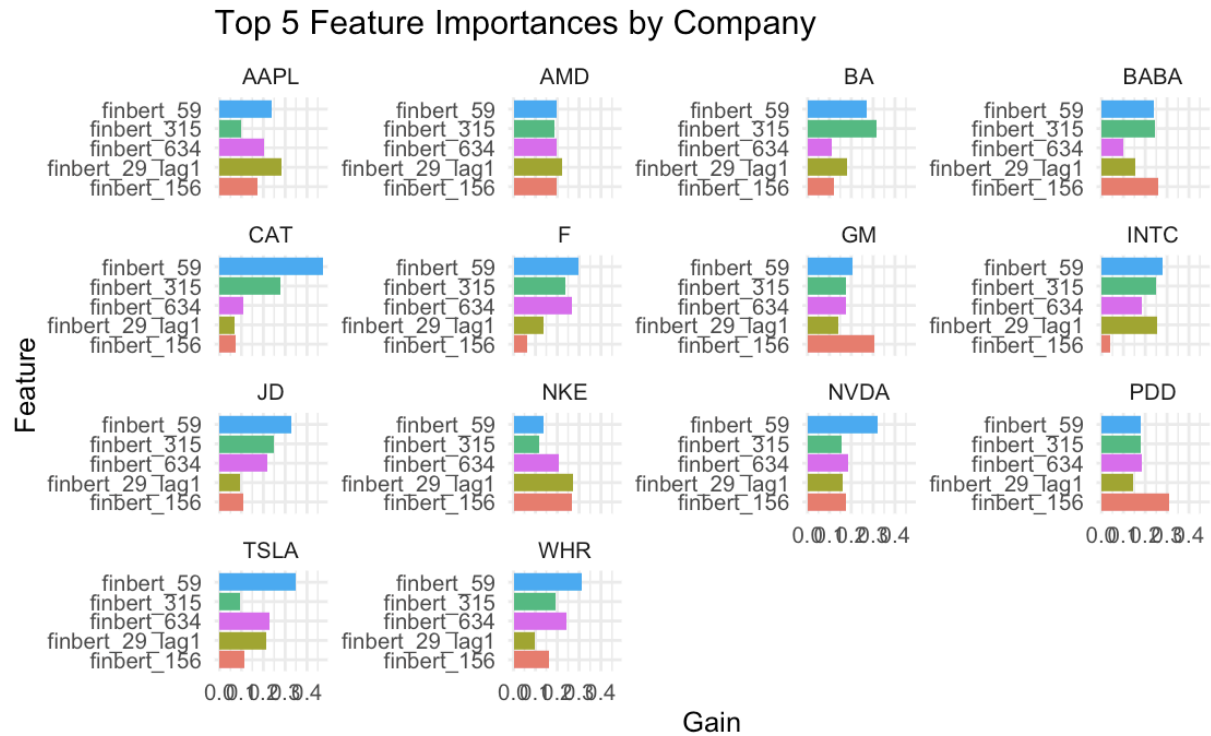
Company-level Lasso Variable Selection



Graph 3: Optimal Lambda(min) by Company (Current & Lagged Embeddings)

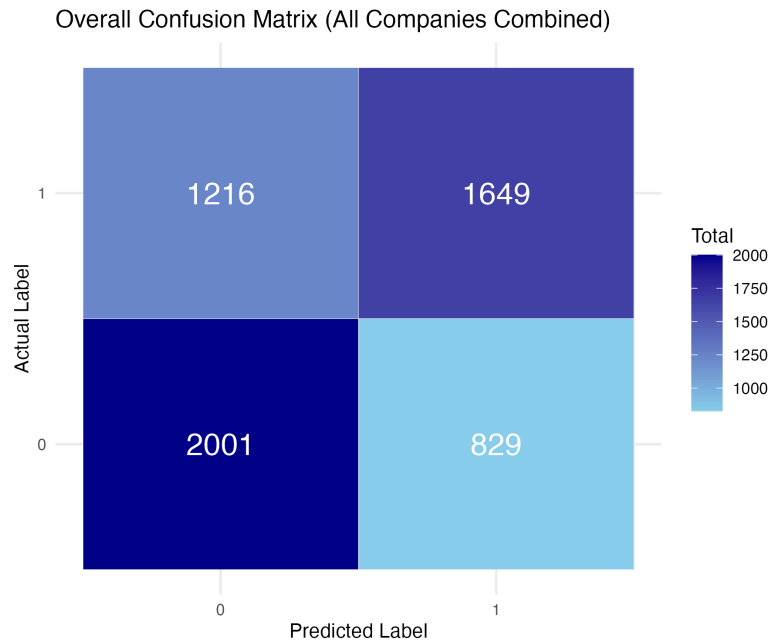
Graph 3 displays the optimal λ (lambda) values obtained through 5-fold cross-validation for each company's LASSO regression model. The regularization parameter λ controls the balance between model complexity and sparsity—higher λ values imply stronger regularization and fewer selected features. Companies such as AMD, NVDA, and PDD exhibit relatively high $\lambda(\text{min})$ values, suggesting their return predictions rely on a more selective subset of FinBERT embedding dimensions and may benefit from stricter regularization to avoid overfitting. Conversely, companies like AAPL, GM, and CAT show lower λ values, indicating that their stock return variations may be influenced by a broader range of semantic features derived from financial text. These differences reflect how the relevance and dispersion of textual signals vary across firms, reinforcing the need for firm-specific modeling when incorporating sentiment-based embeddings into return forecasting.

Company-level XGBoost Results and Accuracy Test



Graph 7: Top 5 Feature Importances by Company (XGBoost Gain Scores)

Graph 7 breaks down feature importance by company, confirming the widespread relevance of `finbert_59`—which appears as the top feature for every firm in the sample. The consistency across firms, combined with variation in the strength and ranking of other features, emphasizes the shared yet firm-specific nature of how sentiment-encoded language influences short-term return dynamics. The presence of lagged features, like `finbert_29_lag1`, reinforces the idea that news effects are not instantaneous and may persist over time, subtly shaping investor behavior across subsequent intervals.



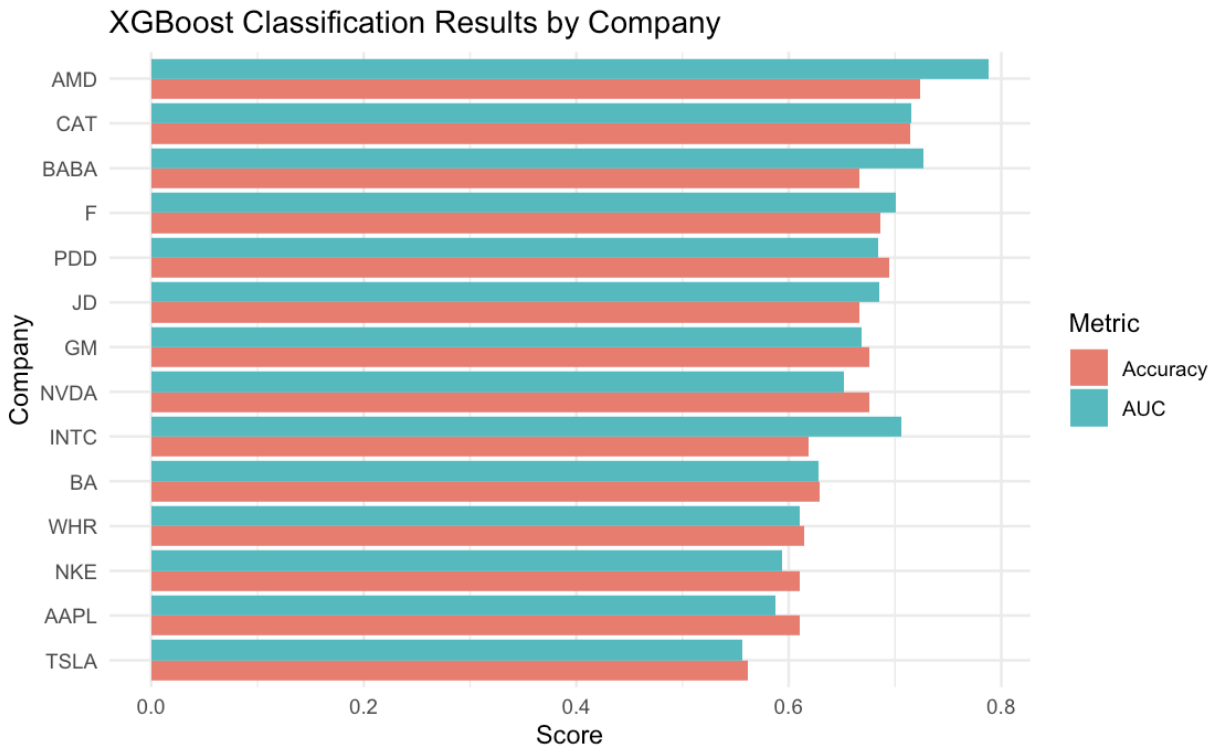
Graph 8. Confusion Matrix of XGBoost Model Across All Companies (Binary Return Direction Classification)

Graph 8 displays the overall confusion matrix for the XGBoost classification model, which was trained to predict short-term stock return directions (positive vs. negative) across all companies. Predictions were made using FinBERT textual embeddings and temporal context (minute-level time blocks), with returns binarized. This distribution indicates that the model has a higher sensitivity to predicting upward movements (higher TP), but struggles more with identifying downward movements (high FN), suggesting a bias toward positive predictions. Despite this, the total number of correct predictions ($TP + TN = 2865$ correct out of 5695 total predictions) shows that the model learns meaningful signals from the financial news data.

Table 1. Company-Level XGBoost Classification Performance of Accuracy and AUC**Scores by Firm**

COMPANY	ACCURACY	AUC
AAPL	0.61	0.588
TSLA	0.562	0.556
INTC	0.619	0.706
AMD	0.724	0.788
NVDA	0.676	0.652
F	0.686	0.701
GM	0.676	0.669
WHR	0.615	0.61
NKE	0.61	0.594
CAT	0.714	0.715
BA	0.629	0.628
BABA	0.667	0.727
JD	0.667	0.685
PDD	0.695	0.684

Table 1 reports the classification performance of XGBoost models for eight companies, using FinBERT embeddings to predict the direction of short-term stock returns. Most companies achieved solid performance, with AMD achieving the highest AUC (0.788) and accuracy (0.724), suggesting that FinBERT-derived features effectively capture news-driven return signals for this stock. CAT and F also performed well, with AUCs of 0.715 and 0.701, respectively. In contrast, TSLA exhibited the weakest results (AUC = 0.556, Accuracy = 0.562), implying that for this firm, sentiment features from financial news may be less predictive or confounded by other factors like volatility or non-news-based drivers. Generally, these results emphasize the firm-specific nature of sentiment predictability and underscore the importance of tailoring models to the responsiveness of each company's stock to textual information.



Graph 9. Accuracy and AUC Scores of XGBoost Classifiers by Company

Graph 9 visualizes Table 1 above. Most companies achieve classification accuracy in the 0.61–0.72 range, with AMD standing out as the top performer (Accuracy = 0.724, AUC = 0.788). This comparative bar chart reinforces the conclusion that the impact of financial news is highly firm-specific.