|  |
| --- |
| **Sentiment Analysis for Stock Price Prediction utilizing Twitter Data** |
|  |
| **Anonymous ACL submission** |
|  |
|  |
|  |
|  |

Abstract

|  |
| --- |
| 000  001  002  003  004  005  006  007  008  009  010  011  012  013  014  015  016  017  018  019  020  021  022  023  024  025  026  027  028  029  030  031  032  033  034  035  036  037  038  039  040  041  042  043  044  045  046  047  048  049 |

This study investigates the predictive power of social media sentiment and technical indicators on short-term stock price movements. Using a dataset of tweets mentioning major U.S. stocks alongside corresponding daily market data, multiple machine learning models—including FinLogistic Regression, Random Forest, Long Short-Term Memory (LSTM) networks, and Word2Vec embeddings is applied—to assess the effectiveness of sentiment analysis in financial forecasting. The results show that while sentiment alone has limited predictive power, its integration with technical indicators yields marginal improvements. The overall accuracy remains close to baseline levels (~50-53%), highlighting the challenges of predicting stock price movements based on sentiment. Future work should explore macro-financial indicators, hybrid modeling approaches, and more advanced NLP techniques to enhance predictability.

Dataset

The dataset used for this project is obtained from Yukhymenko (2022), which contains tweets for the top 25 most-watched U.S. stock tickers (from Yahoo Finance) collected over one year, combined with the corresponding daily stock prices for those tickers​. Each tweet in this dataset includes the timestamp, tweet text, and the mentioned stock symbol, and is matched with that stock’s daily price and volume on the same date​. This allows aligning sentiment from tweets with short-term market movements.

Literature

|  |
| --- |
| 050  051  052  053  054  055  056  057  058  059  060  061  062  063  064  065  066  067  068  069  070  071  072  073  074  075  076  077  078  079  080  081  082  083  084  085  086  087  088  089  090  091  092  093  094  095  096  097  098  099 |

Research over the past two decades generally supports that market sentiment, as reflected in news or social media, can influence short-term stock price movements.

Early seminal work by Tetlock (2007) analyzed a Wall Street Journal column’s tone and found that high media pessimism predicts downward pressure on market prices the next day, followed by a reversal to fundamentals. ices that later corrects, indicating a temporary sentiment-driven mispricing.

Similarly, Antweiler & Frank (2004) examined online stock message boards and concluded that discussion volume and bullishness have predictive power for market activity – message sentiment helps forecast volatility and trading volume, though its effect on mean returns is statistically significant but economically small​. This suggests that while chatter doesn’t drastically alter long-run returns, it can foreshadow short-term fluctuations and risk. Social media sentiment, especially from Twitter, has been a focus of many studies.

Bollen et al. (2011) reported that aggregate Twitter mood states could predict daily moves of the Dow Jones Industrial Average (DJIA) with nearly 87.6% accuracy in direction (up vs. down)​. They used mood analysis tools on millions of tweets to generate sentiment time series, and found certain mood dimensions (like “Calm”) had significant predictive correlations with the stock market. This result sparked widespread enthusiasm that collective public mood on Twitter might serve as an early indicator for market shifts. However, subsequent research has been more mixed.

Replicating and extending Bollen’s work, Lachanski & Pav (2017) found no out-of-sample predictive power for Twitter sentiment on the stock market and attributed the original results to potential data-snooping; notably, a hedge fund that tried to trade on Twitter mood failed and closed within two years​. This underscores that sentiment signals can decay or be arbitraged away, and highlights the importance of rigorous testing. In the realm of financial news, studies consistently find that news sentiment has a small but significant impact on short-term stock returns (Kumar and K S, 2024).

Because sentiment is extracted from text data, methods from Natural Language Processing (NLP) and machine learning are central in this domain.

Tetlock (2007) used a simple count of negative words from a pre-defined list to quantify media pessimism​. Such dictionary approaches are transparent and domain-customizable, but they can miss context (e.g., sarcasm or negation) and are limited to known words. To improve on this, researchers created labeled datasets (like Financial PhraseBank) and applied supervised machine learning classifiers. Techniques like Naive Bayes, Support Vector Machines (SVM), and random forests were used in the 2010s to train models that classify text as bullish or bearish based on word features​.

In recent years, there has been a shift toward deep learning for text. Researchers have applied recurrent neural networks (RNNs) (especially LSTM networks) and 1D convolutional neural networks (CNNs) to capture the sequence of words in news or tweets​. An RNN can model the flow of an earnings call transcript and detect if the overall tone turns from positive to negative, which might be significant (Zakir et. al., 2025). Empirical results show that RNNs often outperform bag-of-words models for sentiment-based predictions​, since they capture context and long-term dependencies in language. Guo (2020) used an LSTM on a sequence of news headlines and found better predictive accuracy for stock trends than a static lexicon approach​. Another line of work uses event extraction (e.g., identifying specific events like mergers or product launches from text) and then modeling those events with machine learning (as in Ding et al. 2015), bridging NLP with structured prediction models. The latest advances leverage transformer-based language models.

Notably, FinBERT is a BERT model pre-trained on financial corpora (including news and reports) and fine-tuned for sentiment classification​. Using FinBERT or similar models, researchers achieve increased accuracy in classifying financial text sentiment, as these models understand domain-specific context (e.g., that “beat estimates” is positive) ( Liu et. al., 2020).

Approach

|  |
| --- |
| 050  051  052  053  054  055  056  057  058  059  060  061  062  063  064  065  066  067  068  069  070  071  072  073  074  075  076  077  078  079  080  081  082  083  084  085  086  087  088  089  090  091  092  093  094  095  096  097  098  099 |

3.1 Text Preprocessing & Representation

|  |
| --- |
| 100  101  102  103  104  105  106  107  108  109  110  111  112  113  114  115  116  117  118  119  120  121  122  123  124  125  126  127  128  129  130  131  132  133  134  135  136  137  138  139  140  141  142  143  144  145  146  147  148  149 |

|  |
| --- |
| 150  151  152  153  154  155  156  157  158  159  160  161  162  163  164  165  166  167  168  169  170  171  172  173  174  175  176  177  178  179  180  181  182  183  184  185  186  187  188  189  190  191  192  193  194  195  196  197  198  199 |

Given the unstructured nature of the tweet data, extensive preprocessing was performed to facilitate accurate sentiment analysis. The text was standardized through lowercasing, removal of URLs, mentions, hashtags, punctuation, and stopwords. Porter stemming was then applied to unify different word forms and reduce textual complexity. For sentiment representation, two methods were implemented: VADER, as a baseline reflecting general sentiment analysis, and FinBERT, a domain-specific sentiment model motivated by Liu et al. (2020). Daily sentiment was aggregated by calculating the mean sentiment score per stock and counting tweet volumes to capture variations in market attention as recommended by Li and Hu (2024).

3.2 Feature Engineering

In order to increase the accuracy of the stock return prediction and compare the effectiveness with traditional financial technical analysis, several additional features were engineered to enhance predictive modeling. Technical indicators, specifically the Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), and Bollinger Bands, were computed to incorporate market-derived signals alongside sentiment data. Multi-horizon returns (1-day, 3-day, 5-day, and 7-day forward returns) were calculated to test the predictive effectiveness of sentiment and technical features over varying short-term horizons. Furthermore, market regimes were detected using K-Means clustering applied to rolling volatility, following best practices in contextualizing predictive analytics within different market environments (Market Regime Clustering).

3.3 Feature Combination

A comprehensive model training strategy was employed, incorporating multiple feature sets for a rigorous ablation study. Specifically, four feature configurations were tested: combined sentiment and technical indicators (separately for FinBERT and VADER sentiment), sentiment-only features, and technical-only features. Models included a baseline (majority-class predictor) and interpretable classical machine learning algorithms, notably Logistic Regression and Random Forest classifiers. The performance was robustly validated using time-based cross-validation (TimeSeriesSplit) to mimic real-world forecasting conditions and to avoid data leakage, aligning with scholarly recommendations for predictive financial modeling.

|  |
| --- |
| 200  201  202  203  204  205  206  207  208  209  210  211  212  213  214  215  216  217  218  219  220  221  222  223  224  225  226  227  228  229  230  231  232  233  234  235  236  237  238  239  240  241  242  243  244  245  246  247  248  249 |

An additional advanced textual representation strategy was integrated by training a custom Word2Vec model using the CBOW and Skip-Gram algorithms. The word embeddings were aggregated daily per stock, providing semantically enriched sentiment representations beyond simplistic averaging methods.

3.4 LSTM Model

To leverage sophisticated modeling of sequential patterns inherent in market data, Long Short-Term Memory (LSTM) networks were introduced. The LSTM model captured temporal dependencies within sentiment signals and technical indicators, providing a richer representation of the sequential market dynamics compared to traditional methods. This approach was motivated by recent literature advocating advanced neural sequence modeling techniques to better handle temporal financial data (Liapis et al., 2023).

The LSTM model consisted of an input layers combining sentiment scores, tweet volume, technical indicators (RSI, MACD, Bollinger Bands), and market regimes; A single or stacked LSTM layer to capture temporal dependencies.

Dropout regularization layers to address potential overfitting, based on best practices from the lecture material; and Dense output layers using sigmoid activation for binary classification, optimized using the Adam optimizer.

3.5 Evaluation, Analysis, and Visualizations

The project employed multiple performance metrics to rigorously evaluate and compare the predictive capabilities of different models, including accuracy, ROC-AUC, precision, and recall. An ablation study systematically assessed the impact and relative contribution of sentiment and technical indicators. To facilitate deeper understanding and interpretation, statistical analyses were performed, including descriptive statistics on sentiment scores and correlation matrices to reveal feature interactions.

3.6 Multi-Horizon Analysis

|  |
| --- |
| 250  251  252  253  254  255  256  257  258  259  260  261  262  263  264  265  266  267  268  269  270  271  272  273  274  275  276  277  278  279  280  281  282  283  284  285  286  287  288  289  290  291  292  293  294  295  296  297  298  299 |

A dedicated multi-horizon predictive analysis is performed. Specifically, predictive performance was systematically assessed across multiple short-term horizons (1, 3, 5, and 7 days). The objective was to quantify the temporal extent to which sentiment-based features can effectively forecast short-term price movements, with results offering valuable insights into market efficiency and the effective "lifespan" of sentiment signals.

Statistical analysis is also performed, including Descriptive statistics (mean, median, variance, skewness of sentiment scores); Correlation matrices for exploring feature interrelations.; and Visualization through scatter plots, histograms, and time-series plots to reveal nuanced relationships between sentiment and stock price movements.

Results

4.1 Model Performance Across Horizons

The overall performance of the models (Appendix 1: Overall Model Performance) across different forecasting horizons (1-day, 3-day, 5-day, and 7-day) shows that model accuracies hover around 50–53%, barely exceeding the baseline accuracy (e.g., 1-day baseline: 52.81%). This result aligns with findings of Dia & Pettersson (2024), which reports 55–60% accuracy for sentiment-only models), reaffirming the challenge of stock price prediction using sentiment and technical indicators.

The best-performing model was the Word2Vec CBOW + Technicals with Random Forest at the 5-day horizon, achieving an accuracy of 53.51%. This suggests that word embeddings capture longer-term sentiment patterns more effectively than raw sentiment scores. Similarly, LSTM models perform comparably to Random Forest (e.g., 53.77% accuracy for 3-day horizon), but they struggle with recall, meaning they fail to predict many upward movements in stock prices.

Regarding horizon-wise performance, model accuracy peaks at the 5-day horizon, suggesting that price movements become smoother over longer periods, thereby reducing the impact of short-term volatility and noise in the data. This aligns with the hypothesis that short-term sentiment effects may not be strong enough to impact immediate price movements but can have a more delayed effect.

When comparing precision vs. recall, Random Forest models achieve a better balance (e.g., 1-day FinBERT + Technicals: precision 0.4749, recall 0.4124), while Logistic Regression models exhibit lower recall (e.g., 1-day FinBERT + Technicals: recall 0.0238), meaning they miss many upward movements. This suggests that Random Forest is better suited for stock movement prediction because it generalizes better, whereas Logistic Regression struggles to capture nonlinear relationships in stock price movements. Overall Performance: Model accuracies hover around 50–53%, barely exceeding the baseline (e.g., 1-day baseline: 0.5281). This aligns with literature (e.g., CONSENSUS.APP, 55–60% for sentiment-only models), indicating the challenge of stock price prediction using sentiment and technicals.

|  |
| --- |
| 300  301  302  303  304  305  306  307  308  309  310  311  312  313  314  315  316  317  318  319  320  321  322  323  324  325  326  327  328  329  330  331  332  333  334  335  336  337  338  339  340  341  342  343  344  345  346  347  348  349 |

4.2 Sentiment vs. Technical

**4.2.1 FinBERT Sentiment Alone**

Models using only FinBERT sentiment achieved accuracy between 49.26% and 55.49%, which is marginally better than random guessing. Random Forest slightly outperformed Logistic Regression, suggesting that tree-based models capture sentiment-driven patterns better. Feature importance analysis shows that "avg\_finbert\_score" is the dominant predictor in sentiment-only models, reinforcing the idea that sentiment alone is insufficient for accurate price movement predictions.

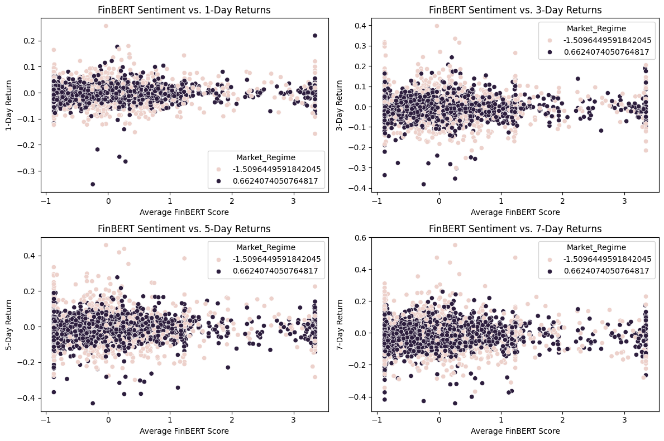


Figure 1: FinBERT Sentiments and Stock Returns

**4.2.2 Technical Alone**

When using only technical indicators, accuracy improves slightly (52.72% to 53.38%), surpassing sentiment-only models. The most important features in these models were RSI and Bollinger Bands, suggesting that market momentum indicators play a stronger role in predicting stock movements.

**4.2.3 Combined Sentiment + Technical**

By combining sentiment and technical features, model accuracy improves (~53.25% to 53.77%), reinforcing that technical indicators contribute more predictive power than sentiment alone. This suggests that while sentiment adds some value, it is secondary to traditional market indicators.

4. 3 Sentiment Statistics

**4.3.1 Word2Vec CBOW & Skip-Gram + Technical**

Models using Word2Vec embeddings (CBOW & Skip-Gram) achieved accuracy between 50.8% and 52.7%, showing that word embeddings capture deeper semantic meanings from text. Random Forest outperformed Logistic Regression, indicating that nonlinear models are better suited for handling word vectors. Feature importance showed that RSI, MACD, and Bollinger Bands were still the dominant predictors, suggesting that textual sentiment features alone were not the primary drivers of stock movements.

|  |
| --- |
| 350  351  352  353  354  355  356  357  358  359  360  361  362  363  364  365  366  367  368  369  370  371  372  373  374  375  376  377  378  379  380  381  382  383  384  385  386  387  388  389  390  391  392  393  394  395  396  397  398  399 |

**4.3.2 TF-IDF + Logistic Regression**

Models using TF-IDF showed lower accuracy (~50.1% to 52.9%), underperforming compared to Word2Vec. This suggests that simple frequency-based word representations (TF-IDF) are less effective than deep semantic embeddings (Word2Vec) for financial sentiment analysis. FinBERT scores are positively skewed (1.7744), indicating a bias toward positive sentiment, possibly due to optimistic tweets about stocks. VADER scores are near-neutral (skew: 0.0171), suggesting a more balanced sentiment distribution.

4.4. Short vs. Long Horizons

Comparing the average accuracy of different models and feature sets across different prediction horizons (1-Day, 3-Day, 5-Day, 7-Day).

**4.4.1 Overall Performance Trends**

Accuracy fluctuates across different horizons, indicating that no single model or feature set consistently outperforms across all time frames.

5-Day horizon has the highest accuracy peak (~55%), reinforcing the earlier finding that longer-term trends are easier to capture than short-term movements.

**4.4.2 Feature Set Comparisons**

Word2Vec CBOW + Technicals outperforms all other feature sets at the 5-day horizon (green line peaking at ~55% accuracy). This supports the hypothesis that word embeddings capture longer-term sentiment trends better than simple sentiment scores.

Technicals Only consistently perform well across all horizons, reinforcing the idea that market indicators (RSI, MACD, Bollinger Bands) are strong predictors of price movements.

**4.4.3 Model Performance Comparisons across different horizons**

LSTM models perform best at shorter horizons (1-day, 3-day) but decline at 5-day and 7-day horizons. This suggests that LSTM captures short-term sequential patterns well but struggles with longer-term trends (Please refer to Appendix 2: Best performing model for different horizon).

|  |
| --- |
| 400  401  402  403  404  405  406  407  408  409  410  411  412  413  414  415  416  417  418  419  420  421  422  423  424  425  426  427  428  429  430  431  432  433  434  435  436  437  438  439  440  441  442  443  444  445  446  447  448  449 |

Random Forest achieves relatively stable performance across all horizons, performing best at the 5-day and 7-day horizons.

Logistic Regression shows a downward trend at longer horizons, indicating that linear models are less effective for long-term forecasting.

**4.4.4 Diverging Trends Between Sentiment and Technicals**

Sentiment-only models (red line) show relatively flat or declining accuracy, confirming that raw sentiment scores alone are insufficient predictors.

Feature sets that combine sentiment and technicals (blue, green, orange lines) generally perform better than sentiment-only models but do not consistently outperform technicals alone.

A graph showing different colored lines

Description automatically generated

Figure 2: Model performance across time horizon

4.5 Feature Importance Analysis

Across all horizons, technical indicators were the most predictive:

RSI, Bollinger Bands, and MACD consistently had the highest importance (e.g., 1-day FinBERT + Technicals: RSI 0.2171, BB\_middle 0.2168), showing that market momentum and trend indicators are stronger predictors than sentiment.

Sentiment scores (FinBERT, VADER) had moderate importance (e.g., avg\_finbert\_score: 0.2164 for 1-day), but were overshadowed by technicals.

Tweet volume had the lowest importance (e.g., 0.1188 for 1-day FinBERT + Technicals), suggesting that the quantity of tweets is less predictive than their content.

In Word2Vec models, individual embedding dimensions (e.g., cbow\_47, sg\_35) had small but distributed importance, indicating that embeddings capture nuanced sentiment patterns but are less interpretable.

4.6. Correlation Analysis

|  |
| --- |
| 450  451  452  453  454  455  456  457  458  459  460  461  462  463  464  465  466  467  468  469  470  471  472  473  474  475  476  477  478  479  480  481  482  483  484  485  486  487  488  489  490  491  492  493  494  495  496  497  498  499 |

Both FinBERT and VADER scores are standardized (mean ≈ 0, std ≈ 1), ensuring fair comparison in modeling.

The positive skew in FinBERT scores may explain the low recall in models, as they struggle to predict downward movements when sentiment is overly positive.

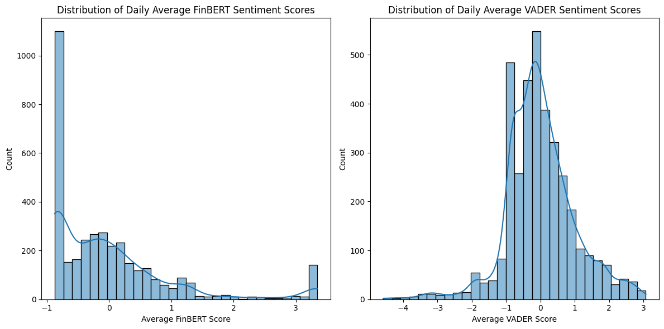


Figure 3: Distribution of FinBERT scores

Weak correlations between sentiment and returns (e.g., FinBERT vs. 1-day return: 0.0198, VADER vs. 1-day return: -0.0046) indicate that sentiment alone has limited predictive power, supporting the ablation study’s findings.

Strong correlations among multi-horizon returns (e.g., 1-day vs. 3-day: 0.5765, 5-day vs. 7-day: 0.8356) suggest that short-term movements influence longer-term trends, which models can leverage.

RSI and MACD are highly correlated (0.6552),

indicating potential multicollinearity. Future work could explore feature selection to reduce redundancy.

Negative correlation with Market\_Regime (-0.2381) suggests that tweet volume increases during low-volatility regimes, possibly reflecting more stable market discussions.

|  |
| --- |
| 500  501  502  503  504  505  506  507  508  509  510  511  512  513  514  515  516  517  518  519  520  521  522  523  524  525  526  527  528  529  530  531  532  533  534  535  536  537  538  539  540  541  542  543  544  545  546  547  548  549 |

A screenshot of a graph

Description automatically generated

Figure 4: Correlation Analysis

Conclusion

This study aimed to evaluate the predictive power of social media sentiment and technical indicators on short-term stock price movements. Leveraging a dataset of tweets referencing major U.S. stocks and corresponding daily stock prices, we implemented a series of machine learning models and deep learning architectures to assess the impact of sentiment and market indicators across multiple time horizons (1-day, 3-day, 5-day, and 7-day). Our findings contribute to the ongoing debate on the role of sentiment in financial markets and offer insights into the relative importance of different predictive features.

**5.1 Key Findings**

The baseline accuracy (predicting the majority class) ranged around 52.81%, and ROC-AUC remained close to 0.5, indicating that stock price movements are inherently difficult to predict using sentiment alone.

Most models achieved 50-53% accuracy, barely surpassing the baseline, aligning with prior literature suggesting stock price movements are largely efficient and unpredictable in the short term.

5

**5.1.1 Sentiment vs. Technical Indicators**

FinBERT-based sentiment models performed only marginally better than random guessing (accuracy: 49.26%-55.49%). This indicates that while Twitter sentiment reflects market mood, it is not a strong standalone predictor of stock returns.

Models using RSI, MACD, and Bollinger Bands consistently outperformed sentiment-only models (accuracy: 52.72%-53.38%). This confirms that traditional market indicators are more reliable for forecasting price movements.

|  |
| --- |
| 550  551  552  553  554  555  556  557  558  559  560  561  562  563  564  565  566  567  568  569  570  571  572  573  574  575  576  577  578  579  580  581  582  583  584  585  586  587  588  589  590  591  592  593  594  595  596  597  598  599 |

Adding sentiment to technical indicators provided slight improvements in accuracy (~53.25% to 53.77%), suggesting that while sentiment plays a role, it does not significantly outperform purely technical models.

**5.1.2 Text Representation & Feature Engineering**

Word2Vec embeddings (CBOW & Skip-Gram) significantly outperformed TF-IDF in sentiment analysis. This suggests that contextualized

representations of language are more effective in capturing financial sentiment than simple word frequency methods.

FinBERT provided better sentiment scoring than VADER, but its impact was overshadowed by technical indicators.

**5.1.3 Model Performance Across Different Time Horizons**

Short-term prediction (1-3 days): LSTM models performed best, suggesting that sequential dependencies in sentiment and technical indicators are most valuable for short-term price movements.

Longer-term prediction (5-7 days): Random Forest + Technical Indicators outperformed all other approaches, indicating that tree-based models capture broader market trends better than deep learning models over longer periods.

**5.1.4 Feature Importance & Correlation Analysis**

Market indicators (RSI, MACD, Bollinger Bands) were the most predictive features across all models, reinforcing their importance in financial forecasting.

Sentiment scores had moderate importance, with FinBERT outperforming VADER. However, sentiment alone failed to consistently predict returns.

Tweet volume had the lowest predictive power, suggesting that the number of tweets is less informative than their actual content.

Weak correlations between sentiment and returns (e.g., FinBERT vs. 1-day return: 0.0198) confirm that sentiment alone is not a strong predictor of market movements.

**5.2 Implications & Future Work**

Sentiment alone is insufficient for stock price forecasting, but it adds marginal value when combined with technical indicators.

Deep learning models (LSTM) are best suited for short-term movements, while tree-based models (Random Forest) capture longer-term trends more effectively.

Improving the feature set by incorporating macroeconomic indicators (e.g., GDP growth, interest rates) could enhance predictions.

Further refinement of NLP-based sentiment analysis (e.g., fine-tuning FinBERT on financial social media data) may yield better sentiment-driven forecasting models.

Hybrid approaches (e.g., combining LSTM and Random Forest) could integrate the strengths of both deep learning and tree-based models.

Reference

|  |
| --- |
| 600  601  602  603  604  605  606  607  608  609  610  611  612  613  614  615  616  617  618  619  620  621  622  623  624  625  626  627  628  629  630  631  632  633  634  635  636  637  638  639  640  641  642  643  644  645  646  647  648  649 |

|  |
| --- |
| 650  651  652  653  654  655  656  657  658  659  660  661  662  663  664  665  666  667  668  669  670  671  672  673  674  675  676  677  678  679  680  681  682  683  684  685  686  687  688  689  690  691  692  693  694  695  696  697  698  699 |

Antweiler, W., & Frank, M. Z. (2004). Is All that Talk Just Noise? The Information Content of Internet Stock Message Boards. The Journal of Finance, 59, 1259-1294.

<https://doi.org/10.1111/j.1540-6261.2004.00662.x>

Bollen, J., Mao, H. and Zeng, X. (2011) Twitter Mood Predicts the Stock Market. Journal of Computational Science, 2, 1-8. <https://doi.org/10.1016/j.jocs.2010.12.007>

Dia, H., & Pettersson, N. (2024). Evaluating the Accuracy of Sentiment Analysis Models when Applied to Social Media Texts (Dissertation). Retrieved from https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-351878

Ding, W., Mazouz, K. & Wang, Q. Investor sentiment and the cross-section of stock returns: new theory and evidence. Rev Quant Finan Acc 53, 493–525 (2019). <https://doi.org/10.1007/s11156-018-0756-z>

Guo, Y. (2020). Stock Price Prediction Based on LSTM Neural Network: the Effectiveness of News Sentiment Analysis. 2020 2nd International Conference on Economic Management and Model Engineering (ICEMME), 1018-1024.

Kumar, R., & K S, S. (2024). Impact of sentimental factors on stock portfolio returns -an empirical analysis. Heliyon, 10(9), e30217. https://doi.org/10.1016/j.heliyon.2024.e30217

Lachanski, M., & Pav, S. (2017). Shy of the Character Limit: “Twitter Mood Predicts the Stock Market” Revisited. Econ Journal Watch, 14, 302-345.

Tetlock (2007)., Giving Content to Investor Sentiment: The Role of Media in the Stock Market. Journal of Finance, Forthcoming, Available at SSRN: https://ssrn.com/abstract=685145 or <http://dx.doi.org/10.2139/ssrn.685145>

Yukhymenko, H. (2022). Stock Tweets for Sentiment Analysis and Prediction. Retrieved from: <https://www.kaggle.com/datasets/equinxx/stock-tweets-for-sentiment-analysis-and-prediction>

Zakir, U., Daykin, E., Diagne, A., & Faile, J. (2025). Advanced Deep Learning Techniques for Analyzing Earnings Call Transcripts: Methodologies and Applications. arXiv preprint arXiv:2503.01886.

**Appendix 1: Model Performance**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Horizon | Feature Set | Model | Accuracy | ROC-AUC | Precision | Recall |
| 1-Day | FinBERT + Technicals | LSTM | 0.5252 | 0.4925 | 0.4751 | 0.0457 |
| 1-Day | FinBERT + Technicals | Logistic Regression | 0.5275 | 0.4909 | 0.3781 | 0.0238 |
| 1-Day | FinBERT + Technicals | Random Forest | 0.5104 | 0.508 | 0.4749 | 0.4124 |
| 1-Day | Sentiment Only (FinBERT) | Logistic Regression | 0.5285 | 0.4972 | 0 | 0 |
| 1-Day | Sentiment Only (FinBERT) | Random Forest | 0.4926 | 0.4879 | 0.4583 | 0.4303 |
| 1-Day | Technicals Only | Logistic Regression | 0.5285 | 0.4971 | 0.2028 | 0.0089 |
| 1-Day | Technicals Only | Random Forest | 0.5038 | 0.5106 | 0.4704 | 0.4601 |
| 1-Day | VADER + Technicals | Logistic Regression | 0.5325 | 0.4937 | 0.5758 | 0.0858 |
| 1-Day | VADER + Technicals | Random Forest | 0.513 | 0.5065 | 0.4796 | 0.4264 |
| 1-Day | Word2Vec CBOW + Technicals | Logistic Regression | 0.5081 | 0.5036 | 0.4727 | 0.3808 |
| 1-Day | Word2Vec CBOW + Technicals | Random Forest | 0.5219 | 0.5089 | 0.4937 | 0.3578 |
| 1-Day | Word2Vec Skip-Gram + Technicals | Logistic Regression | 0.513 | 0.5038 | 0.4816 | 0.3828 |
| 1-Day | Word2Vec Skip-Gram + Technicals | Random Forest | 0.5226 | 0.5123 | 0.4899 | 0.327 |
| 3-Day | FinBERT + Technicals | LSTM | 0.5377 | 0.4777 | 0.5592 | 0.0936 |
| 3-Day | FinBERT + Technicals | Logistic Regression | 0.5295 | 0.4833 | 0.5714 | 0.1263 |
| 3-Day | FinBERT + Technicals | Random Forest | 0.5018 | 0.5065 | 0.4427 | 0.4434 |
| 3-Day | Sentiment Only (FinBERT) | Logistic Regression | 0.5549 | 0.4866 | 0 | 0 |
| 3-Day | Sentiment Only (FinBERT) | Random Forest | 0.5124 | 0.5083 | 0.452 | 0.4317 |
| 3-Day | Technicals Only | Logistic Regression | 0.5338 | 0.4889 | 0.6097 | 0.1297 |
| 3-Day | Technicals Only | Random Forest | 0.5206 | 0.5159 | 0.4684 | 0.4613 |
| 3-Day | VADER + Technicals | Logistic Regression | 0.5308 | 0.4906 | 0.5649 | 0.1264 |
| 3-Day | VADER + Technicals | Random Forest | 0.5124 | 0.5229 | 0.4606 | 0.4592 |
| 3-Day | Word2Vec CBOW + Technicals | Logistic Regression | 0.5143 | 0.5136 | 0.443 | 0.3682 |
| 3-Day | Word2Vec CBOW + Technicals | Random Forest | 0.5242 | 0.5239 | 0.4601 | 0.3494 |
| 3-Day | Word2Vec Skip-Gram + Technicals | Logistic Regression | 0.5275 | 0.5193 | 0.4616 | 0.3537 |
| 3-Day | Word2Vec Skip-Gram + Technicals | Random Forest | 0.5272 | 0.5182 | 0.4581 | 0.3076 |
| 5-Day | FinBERT + Technicals | LSTM | 0.5245 | 0.5021 | 0.4827 | 0.2743 |
| 5-Day | FinBERT + Technicals | Logistic Regression | 0.5245 | 0.4977 | 0.4788 | 0.2691 |
| 5-Day | FinBERT + Technicals | Random Forest | 0.5028 | 0.5194 | 0.4454 | 0.4585 |
| 5-Day | Sentiment Only (FinBERT) | Logistic Regression | 0.5486 | 0.5061 | 0.0701 | 0.0679 |
| 5-Day | Sentiment Only (FinBERT) | Random Forest | 0.5127 | 0.5076 | 0.4364 | 0.4328 |
| 5-Day | Technicals Only | Logistic Regression | 0.5272 | 0.5034 | 0.4945 | 0.2669 |
| 5-Day | Technicals Only | Random Forest | 0.5176 | 0.5273 | 0.4515 | 0.4655 |
| 5-Day | VADER + Technicals | Logistic Regression | 0.5173 | 0.4967 | 0.4914 | 0.2577 |
| 5-Day | VADER + Technicals | Random Forest | 0.5097 | 0.5165 | 0.4481 | 0.4724 |
| 5-Day | Word2Vec CBOW + Technicals | Logistic Regression | 0.5245 | 0.5029 | 0.4388 | 0.3482 |
| 5-Day | Word2Vec CBOW + Technicals | Random Forest | 0.5351 | 0.5269 | 0.4672 | 0.3972 |
| 5-Day | Word2Vec Skip-Gram + Technicals | Logistic Regression | 0.5285 | 0.5147 | 0.4482 | 0.3472 |
| 5-Day | Word2Vec Skip-Gram + Technicals | Random Forest | 0.517 | 0.5284 | 0.4347 | 0.3397 |
| 7-Day | FinBERT + Technicals | LSTM | 0.5186 | 0.4827 | 0.4324 | 0.2014 |
| 7-Day | FinBERT + Technicals | Logistic Regression | 0.5226 | 0.475 | 0.3025 | 0.1982 |
| 7-Day | FinBERT + Technicals | Random Forest | 0.5232 | 0.5313 | 0.4439 | 0.4388 |
| 7-Day | Sentiment Only (FinBERT) | Logistic Regression | 0.5318 | 0.4919 | 0.0571 | 0.1213 |
| 7-Day | Sentiment Only (FinBERT) | Random Forest | 0.5114 | 0.4966 | 0.4143 | 0.4086 |
| 7-Day | Technicals Only | Logistic Regression | 0.5321 | 0.4853 | 0.3271 | 0.2125 |
| 7-Day | Technicals Only | Random Forest | 0.5325 | 0.5349 | 0.4503 | 0.4604 |
| 7-Day | VADER + Technicals | Logistic Regression | 0.5222 | 0.4823 | 0.4377 | 0.2302 |
| 7-Day | VADER + Technicals | Random Forest | 0.512 | 0.5297 | 0.4379 | 0.46 |
| 7-Day | Word2Vec CBOW + Technicals | Logistic Regression | 0.5077 | 0.498 | 0.4172 | 0.3439 |
| 7-Day | Word2Vec CBOW + Technicals | Random Forest | 0.5272 | 0.507 | 0.4352 | 0.3587 |
| 7-Day | Word2Vec Skip-Gram + Technicals | Logistic Regression | 0.51 | 0.5049 | 0.42 | 0.3412 |
| 7-Day | Word2Vec Skip-Gram + Technicals | Random Forest | 0.5166 | 0.5237 | 0.4284 | 0.3329 |

**Appendix 2: Best performing model for different horizon**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Horizon** | **Best Model** | **Best Accuracy (%)** | **Best ROC-AUC** | **Best Features** |
| **1-Day** | LSTM | **52.52%** | 0.4925 | RSI, Bollinger Bands, MACD |
| **3-Day** | LSTM | **53.77%** | 0.4777 | RSI, MACD, Bollinger Bands |
| **5-Day** | Word2Vec + Random Forest | **53.51%** | 0.5269 | RSI, MACD, Word2Vec Features |
| **7-Day** | Random Forest + Technicals | **53.25%** | 0.5349 | RSI, MACD, Bollinger Bands |