Sentiment-Driven Markets: Enhancing ETF Price Prediction with News and Tweet Analysis

JAKUB Říha

JOSHUA POON

GIANCARLO CAPICOTTO

University of Toronto
Department of Economics
ECO482: Machine Learning Applications in Macroeconomic Finance

April 6th 2025

ABSTRACT

This study explores whether integrating sentiment data from Donald Trump's tweets and major financial news headlines can improve short-term predictions of SPY ETF price direction. Using a dataset of 754 daily observations from 2017–2019, we construct a range of financial and sentiment-based features and train a set of machine learning classifiers with varying features, including logistic regression, KNN, random forests, and XGBoost. Models are evaluated using time-series cross-validation and tested out-of-sample based on classification accuracy, AUROC, cumulative strategy returns. The best-performing model, a shallow random forest using Trump VADER sentiment, achieved a 27% return and a Sharpe Ratio of 3.85 in the test period, outperforming more complex models and the market benchmark. However, many models failed to translate strong in-sample performance into out-of-sample profitability, highlighting limitations tied to overfitting, small sample size, and classification-based objectives.

Replication Link: https://github.com/Jakub-Riha/ECO482.git

I. Introduction

In recent months, stock markets have experienced significant single-day fluctuations in response to political developments and major news events. On November 5th, 2024, Donald Trump was declared the winner of the U.S. presidential election. By the next trading day, the S&P 500 had surged 2.5%. Just a few months later, on April 2nd, 2025, a day Trump labelled "Liberation Day", he announced a sweeping tariff plan, prompting the market to contract by 11.2% over the following seven days (Yahoo Finance, 2025). These examples reflect how rapidly markets can react to political announcements and sentiment-laden news. They also raise an important question: can models that incorporate this type of sentiment systematically to forecast price direction?

According to the Efficient Market Hypothesis (EMH), financial markets are informationally efficient: asset prices reflect all publicly available information, and arbitrage ensures that no investor can consistently achieve abnormal returns. However, recent developments in behavioural finance challenge this assumption by highlighting the role of investor sentiment, media framing, and cognitive biases in driving price dynamics. The increasing availability of social media and news data, combined with advances in natural language processing (NLP), offers new opportunities to test these competing views empirically.

This paper investigates whether the predictive accuracy of market price movement models can be improved by integrating historical price data with sentiment analysis derived from Trump tweets and supplimented by major financial news headlines (NYT, WSJ, WP). Our main research question is: *Can we improve the accuracy of market price movement predictions by integrating historical price data with sentiment analysis of Trump tweets and news headlines?* Our hypothesis, informed by the literature, is that sentiment-enhanced models will outperform purely technical models in forecasting short-term price direction, particularly during politically charged periods. We also expect that Trump's tweets, which are often direct and emotionally charged, will yield stronger predictive signals than more neutral or institutional news coverage.

This work makes four contributions. First, it provides empirical evidence on the predictive value of political communication, specifically Trump's tweets, for short-term market movements. Second, it offers insights into which types of variables are most informative for forecasting. Third, by using ML-generated trading signals, we indirectly test market efficiency, evaluating whether such models can consistently outperform a passive buy-and-hold SPY strategy. Lastly, the paper contributes to behavioural finance by showing how investor sentiment, especially during moments of heightened political tension, can introduce exploitable inefficiencies.

Our research builds on a growing literature in machine learning and finance. For instance, Karabulut (2013) uses Facebook's Gross National Happiness Index to predict market activity, while Tetlock (2007) focuses on the tone of a single Wall Street Journal column to capture media sentiment. More closely related to our work, Bollen et al. (2011) find that Twitter mood can predict changes in the Dow Jones Industrial Average. Unlike these studies, our work combines multiple sentiment sources, including Trump's personal Twitter feed, with traditional financial indicators and applies multiple machine learning models to test comparative forecasting power. Furthermore, while prior research has often emphasised long-term trends or general market indices, we focus specifically on short-term gains in SPY ETFs, making our results directly relevant to active trading strategies.

II. Data & Methodology

Our dataset spans 754 daily observations from 13/01/2017 to 30/12/2019, encompassing three years of Donald Trump's first presidency term. We purposely leave out the year of 2020 due to market noise introduced by the COVID-19 pandemic and subsuquent lock-downs. We sourced historical price data from Yahoo Finance using the yfinancePython API and constructed a set of widely used technical indicators, including returns, volatility (VIX), Moving Average (MA), Strength (RSI), Moving Average Converge Divergence (MACD), Volume, and Ichimoku Cloud. We use Trump's historical tweets, sourced from a publicly available dataset (Reese, 2020), scraped using the GetOldTweets tool by Henrique (2015). Newspaper headlines were collected via TDM Studio's licensed API.

Since the goal of the project is to predict daily price movements, we define our “Target” as a binary dummy variable that takes 1 if next day’s closing price is higher than today's closing price, and 0 if next day’s closing price is lower than today’s closing price:

$$\text{Target}_t = \begin{cases} 1, & \text{if } P_{t+1}^{\text{close}} > P_t^{\text{close}} \\ 0, & \text{if } P_{t+1}^{\text{close}} \leq P_t^{\text{close}} \end{cases}$$

To ensure fair model training, all features were standardised using StandardScaler, preventing features with large magnitudes, like moving averages, from dominating smaller-scale features like sentiment in the distance or regularisation-based models like KNN or regression. *Figure I* presents the correlation matrix of key technical indicators used in the model. Notably, Ichimoku Base, MA50, and MA20 display extremely high correlations (~0.99), reflecting their shared role in capturing medium- to long-term price trends. Despite their redundancy, we retained all three indicators, as preliminary testing showed improved model accuracy when each was included, likely due to subtle differences in the way they respond to price movements. *Figure II* complements this by illustrating the evolution of SPY prices over the sample period, highlighting several major market events and cycles, including extended rallies and corrections, offering important context for interpreting model performance and feature behaviour.

FIGURE I

CORRELATION PLOT OF KEY TECHNICAL VARIABLES

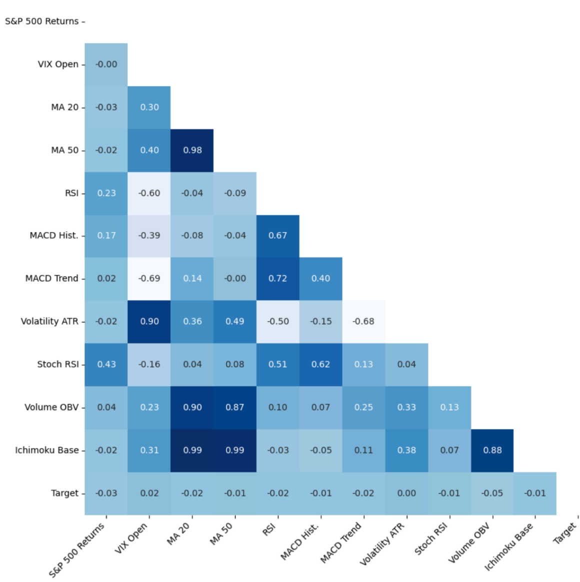
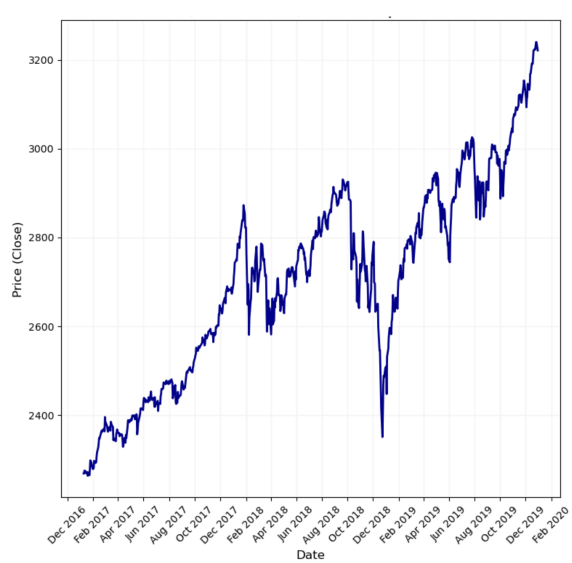


FIGURE II

SPY PRICE MOVEMENT OVER SAMPLE PERIOD



To capture temporal dependencies, we incorporated one-day and three-day lagged versions of each financial indicator. This allowed the models to learn from recent price and momentum trends. After merging all technical and sentiment data, we backfilled missing values and removed the S&P 500 close price, as it would not be observable at prediction time.[1]

For sentiment analysis, we began by cleaning, tokenising, and lemmatising all textual inputs. Given the informal and emotionally charged nature of Trump's tweets, we applied the VADER lexicon, which is optimised for social media sentiment detection. For the news headlines, which are sometimes sarcastic, or ambiguous in tone, we also applied SentiWordNet (SWN) to complement VADER and capture additional linguistic details. Although VADER is not typically used for formal headlines, we found it surprisingly effective in our context, offering predictive value in cross-validation results. By using both lexicons, we aimed to reduce measurement error and broaden sentiment coverage. In line with Gjerstad et al. (2020), we also introduced topic-based features for Trump tweets. Using a Bag-of-Words model and Latent Dirichlet Allocation (LDA), we classified tweets into daily topic labels before merging with our other sentiment features. This allowed us to test the marginal predictive value of each outlet's sentiment in isolation.

The full feature space, comprising technical indicators, sentiment scores, topic frequencies, and lagged versions, was used to train a range of classification models to predict the daily direction of the S&P 500 index. Specifically, we trained eleven models, including logistic regression (with both L1 and L2 penalties), k-nearest neighbors (KNN), random forest classifiers with varying depths, and gradient boosting methods such as XGBoost.

We split the dataset chronologically, using the first 80% of the sample as a training set and the remaining 20% as a test set. Within the training period, we employed 4-fold time-series cross-validation to evaluate each model-feature set combination. For every feature set, we retained the model with the highest cumulative PnL (Profit and Loss) from a simulated trading strategy.

[1] Full Summary Statistics of all features are shown in Appendix I and II

To simulate trades, we assumed that if the predicted direction for day $t+1$ was up, the strategy would go long SPY that day, and go short otherwise. For simplicity, we assumed no trading costs or transaction fees. The cumulative strategy return over the test set is computed as:

$$\text{Strategy PnL} = \sum \left(\text{Position}_t \cdot R_{t+1}\right), \quad \text{where} \quad \text{Position}_t = \begin{cases} +1 & \text{if model predicts up at } t \\ -1 & \text{if model predicts down at } t \end{cases}$$

Model performance was assessed using accuracy, precision, recall, F1 score, AUROC, Sharpe Ratio, and most importantly, the economic PnL of the trading strategy. The Sharpe Ratio was calculated as an annualised measure with 252 trading days:

$$\text{Sharpe Ratio} = \frac{\mu_r}{\sigma_r} \times \sqrt{252}, \quad \text{where } \mu_r \text{ is the mean daily return and } \sigma_r \text{ is the standard deviation of daily returns.}$$

We looped through all our defined models for each feature set, picking only the model with the highest PnL in the training sample and saved it as the "Best Model" for out of sample testing. We found that while cross-validation PnL was often high for several models, test performance frequently regressed toward the mean, highlighting the limited sample size and risk of overfitting.

III. Results

To evaluate the robustness and economic significance of each feature set, we began by analysing out-of-sample results from the best-performing model for each set, as determined by cross-validated strategy PnL. The test results were assessed using the same suite of metrics: accuracy, precision, recall, F1 score, AUROC, Sharpe Ratio, and cumulative strategy PnL. A focused summary for the four most illustrative models, Trump VADER, Trump VADER (Lags), Financial + Trump Themes, and Everything (All Features), is shown in *Table I*.[2]

TABLE I

TRAINING AND TEST PERFORMANCE OF SELECTED MODELS

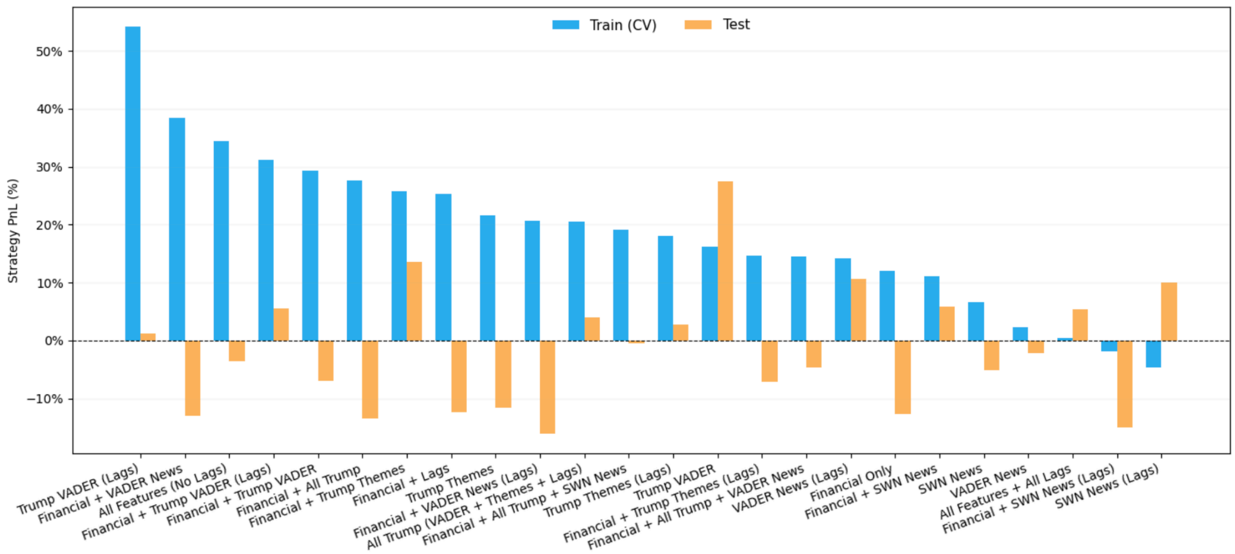
	Model Type	Training Accuracy	Test Accuracy	Training AUROC	Test AUROC	Training PnL	Test PnL	Training Sharpe Ratio	Test Sharpe Ratio
Trump VADER	RF (depth=4)	0.51	0.61	0.52	0.61	16.3%	27.4%	0.62	3.85
Trump VADER (Lags)	RF (depth=8)	0.56	0.55	0.53	0.54	54.1%	1.2%	2.09	0.16
Financial + Trump Themes	KNN (k=3)	0.53	0.55	0.56	0.56	25.7%	13.6%	0.99	1.87
Everything (All Features)	RF (depth=12)	0.51	0.52	0.49	0.55	0.4%	5.4%	0.02	0.74

Overall, performance on the test set was weaker than in training, with many feature sets experiencing a marked drop in both PnL and Sharpe Ratio relative to their cross-validation performance. Despite this, a few models maintained their edge. Notably, **Trump VADER** emerged as the best-performing model in the out-of-sample period. The model achieved a test Sharpe Ratio of 3.85 and a cumulative PnL of 27%, significantly outperforming other strategies. Interestingly, this model underperformed in training (Sharpe Ratio = 0.62; PnL = 16%), suggesting that its predictive power generalised well to unseen data. The training and test sample performance of all models is illustarted below in *figure III*:

[2] A complete table of all test results and metrics can be found in *Appendix III and IV*

TABLE III

PnL (%) OF SELECTED MODELS IN TRAINING AND TESTING SAMPLES



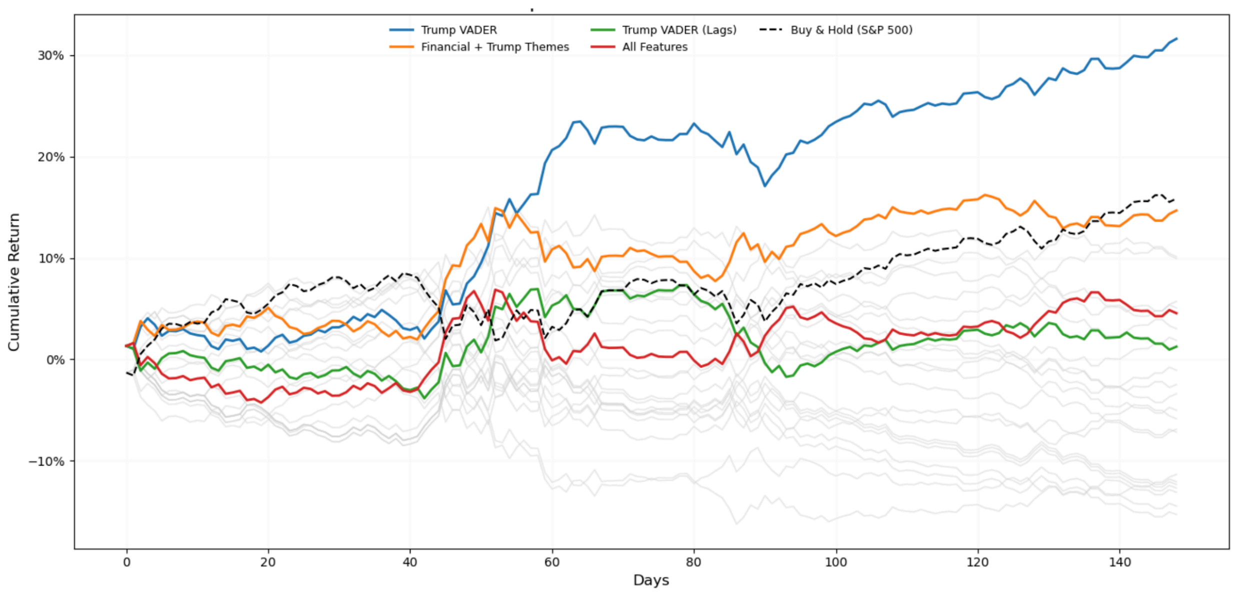
Furthermore, **Trump VADER (Lags),** which had the highest training Sharpe Ratio (2.09) and training PnL (54%), saw a large drop in the test set, where its Sharpe Ratio fell to 0.16 and cumulative PnL to 0.01. This suggests that although lagged sentiment may capture patterns well in-sample, it risks overfitting without sufficient regularisation or validation. Additionally, both Trump VADER models were Random Forest based, with the lagged specification using a maximum depth of 8 compared to the non-lagged model which used a depth of 4. This suggests the additional depth may have been detrimental in generalising the training performance to unseen data.

The **Financial + Trump Themes** model, trained using KNN (k=3), delivered consistently solid results across both periods, with a training Sharpe Ratio of 0.99 and test Sharpe Ratio of 1.87, alongside consistent PnLs of 26% and 14%, respectively. These results show the robustness of combining financial and political text features in models that rely on local distance metrics. Finally, the **Everything (All Features)** model, despite its broad information set, achieved weak performance across both periods. Its test PnL was just 5%, and its Sharpe Ratio remained below 1 in both train and test (0.02 and 0.74, respectively), likely due to noise from redundant variables, or possibly the curve of multidimensionality, as the dataset sufferes from a small sample size (n=754).

Figure IV illustrates the cumulative returns of all strategies on the test set. While most strategies hovered near the market benchmark, **Trump VADER (Lags)** and **Financial + Trump Themes** notably outperformed the S&P 500 Buy & Hold strategy. The steep divergence of the top curve after the 60-day mark confirms the predictive edge of lagged Trump sentiment variables when used with shallower decision trees for greater generalisibility.

FIGURE IV

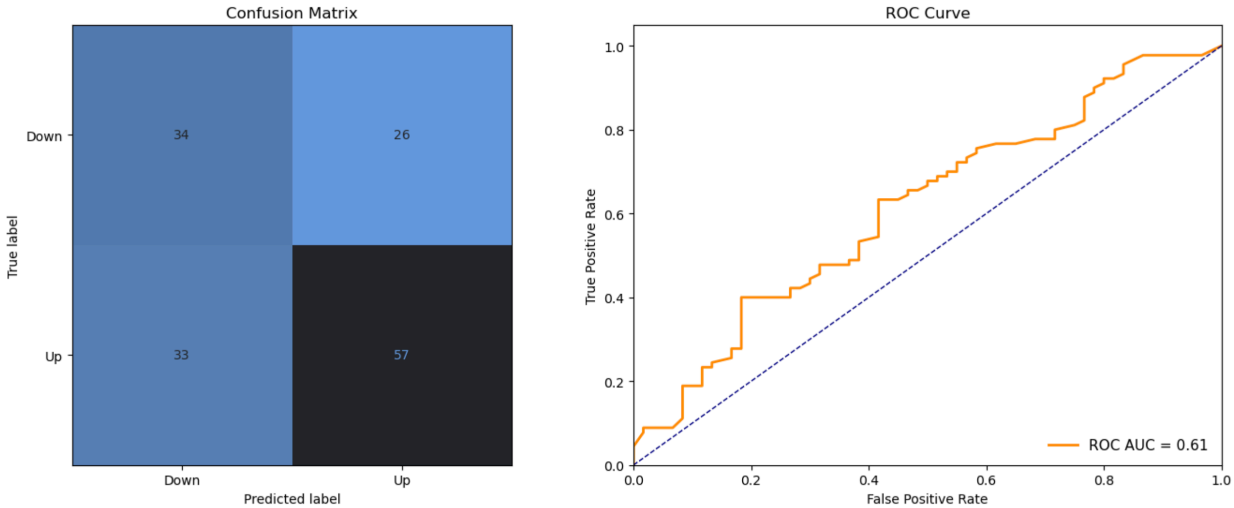
CUMULATIVE RETURNS OF SELECTED MODELS IN TEST SAMPLE



To assess the predictive quality of the top-performing model, we analysed its classification metrics in more depth. *Figure V* below shows the confusion matrix and ROC curve for the **Trump VADER** model on the test set, the overall best performing model. The model achieved a balanced ability to identify both upward and downward market movements, with 57 correct upward classifications out of 90 and 34 correct downward classifications out of 60. Misclassifications occurred in both directions, but the distribution indicates a moderate ability to separate classes in a realistic, noisy financial environment.

FIGURE V

TEST PERFORMANCE OF THE TRUMP VADER MODEL



The ROC curve highlights this performance more formally. The Trump VADER model achieved an out-of-sample AUROC of 0.61, modestly above random guessing, indicating that it has genuine predictive signal, albeit noisy and inconsistent, as is common in financial return prediction. The model's moderate AUROC is complemented by a high Sharpe Ratio, suggesting that while its classification margins may not always be strong, its directional signals translate effectively into profitable trading outcomes. In financial matkets, this distinction is crucial. A model may show middling classification metrics yet perform well economically if the correct predictions coincide with high-return days. Conversely, a model with high accuracy but poor timing can underperform.

However, as this was a classification problem, the models were not trained on the magnitude of the price increase, instead, they were trained on a binary target variable indicating whether the price went up or down in general. For example, many models used the Random Forest classifier to predict the price movements. For classification tasks, tree-based models optimise a simple objective: minimising the number of misclassified samples, as shown below:

$$\min \sum_{i=1}^{n} [y_i \neq \hat{y}_i]$$

where y_i is the true label, and $\hat{y}_i$ is the predicted label. This formulation prioritises accuracy in class prediction rather than the economic impact of those predictions. As a result, a model might optimise for class balance but miss larger market moves that would have been more profitable.

We suspect this minimisation design is the reason behind such dramatic differences between the training and test performance of many models. While some models exhibited high high performance and scored well in many classification metrics on the training sample, once exposed to the test sample, they were either unable to replicate their training PnL, or made a loss overall. For example, the Trump VADER Lag model performed exceptionally well in the training data (54.1% PnL), but it fell short of its training success in the test data (1.2% PnL). However, in both cases the model exhibited relatively strong and consistent performance on accuracy (0.56 vs 0.55), and even overperformed its training AUROC (0.53 vs 0.54). This implies that while the model's objective performance in terms of predicting the price going up vs down remained relatively stable, its ability to translate those predictions into profitable trades diminished, likely due to misaligned signal timing or a lack of sensitivity to the magnitude of returns on those days.

IV. Conclusion

This paper set out to evaluate whether integrating sentiment analysis from Trump tweets and major financial news headlines could enhance the predictive accuracy of machine learning models forecasting daily SPY ETF price direction. Drawing on a rich dataset of technical indicators, sentiment scores, topic frequencies, and their respective lags, we trained eleven classification models using a rolling time-series cross-validation framework. Models were evaluated not only by traditional classification metrics such as accuracy and AUROC, but more importantly by the financial viability of their predictions, measured via cumulative PnL and the Sharpe Ratio.

Our findings indicate that while many models performed well in-sample, only a few maintained strong out-of-sample results. The Trump VADER model, trained using a shallow Random Forest, emerged as the most effective, achieving a test Sharpe Ratio of 3.85 and a cumulative return of 27%, outperforming both the buy-and-hold benchmark and more complex, feature-rich models. Interestingly, this model exhibited modest training performance, suggesting that its predictive signals were less prone to overfitting and more robust in generalising to unseen data.

However, several limitations mitigate these results. First, the dataset spans a relatively short time frame (n=754), which may constrain model generalisability and increase susceptibility to noise. Second, all models were trained as binary classifiers, optimising for directional accuracy rather than financial magnitude. As such, many models may have predicted correctly on average but failed to capture large, high-return days crucial for trading profitability. Finally, the minimisation objective of classification models may prioritise balanced prediction over economic value, which explains why models with decent accuracy still struggled to generate stable trading returns.

Future research could expand this work by incorporating regression-based models that directly predict return magnitudes, enabling better alignment between predictive accuracy and trading profitability. Moreover, more advanced sentiment embedding techniques (BERT or FinBERT) could be employed to extract deeper textual signals from both tweets and headlines. Finally, applying the framework across different time periods and asset classes would offer further insights into the generalisability of sentiment-driven prediction strategies.

IV. References

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

Henrique, J. (2015). *GetOldTweets-python* [Python software]. GitHub. https://github.com/Jefferson-Henrique/GetOldTweets-python

Karabulut, Y. (2013). *Can Facebook predict stock market activity?* Goethe University Frankfurt. https://doi.org/10.2139/ssrn.1910338

Reese, A. (2020). *Trump Tweets* [Data set]. Kaggle. https://www.kaggle.com/datasets/austinreese/trump-tweets

TDM Studio. (n.d.). *News API platform*. ProQuest. https://www.proquest.com/tdmstudio

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance, 62*(3), 1139–1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

V. Appendix

APPENDIX I

SUMMARY STATISTICS OF TECHNICAL INDICATORS

	Observations	Mean	Standard Deviation	Min	Q1	Q2	Q3	Max
Target	754	1	0	0	0	1	1	1
S&P500 Returns	754	0.00	0.01	-0.04	-0.00	0.00	0.01	0.05
VIX	754	14.40	4.10	9.00	11.60	13.20	16.20	37.30
MA20	754	2,691.20	222.60	2,252.70	2,496.40	2,720.20	2,866.40	3,179.90
MA50	754	2,672.80	219.90	2,197.20	2,474.80	2,705.00	2,858.20	3,120.20
RSI	754	59.00	16.50	4.00	47.60	60.50	71.00	94.70
MACD Historical	754	0.10	7.00	-30.60	-2.80	0.40	4.00	21.90
MACD Trend	754	0.86	1.83	-6.83	0.02	1.22	2.09	4.32
ATR	754	2.36	1.21	0.00	1.42	1.98	3.23	6.52
stochRSI	754	0.58	0.36	0.00	0.25	0.62	0.93	1.00
Volume (1000s)	754	1,848,200	809,826	131,122	1,055,902	2,191,579	2,479,921	3,379,715
Ichimoku	754	240.00	22.82	197.20	220.59	240.66	258.04	290.90

APPENDIX II

SUMMARY STATISTICS OF SENTIMENT INDICATORS

	Observations	Mean	Standard Deviation	Min	Q1	Q2	Q3	Max
NYT Vader Score	754	-0.01	0.06	-0.22	-0.05	-0.01	0.03	0.2
NYT SWN Score	754	0.01	0.02	-0.05	0	0.01	0.02	0.07
WP Vader Score	754	-0.02	0.09	-0.32	-0.07	-0.02	0.04	0.29
WP SWN Score	754	0.01	0.02	-0.07	0	0.01	0.02	0.15
WSJ Vader Score	754	0.06	0.07	-0.18	0.02	0.07	0.11	0.23
WSJ SWN Score	754	0.01	0.01	-0.02	0	0.01	0.01	0.03
Trump Vader Score	754	0.6	0.68	-1	0.65	0.96	0.99	1
Patriotic Messaging	754	0.11	0.28	0	0	0	0	1
Immigration & Media	754	0.29	0.37	0	0	0.01	0.55	1
Policy & Trade	754	0.07	0.23	0	0	0	0	1
Foreign Affairs	754	0.03	0.14	0	0	0	0	0.99
Elections & Fake News	754	0.06	0.21	0	0	0	0	0.99
Security & Foreign Influence	754	0.03	0.15	0	0	0	0	0.99
Domestic Events & Disaster	754	0.02	0.11	0	0	0	0	0.98
State Politics & Crime	754	0.07	0.22	0	0	0	0	1
Global & Economy	754	0.05	0.18	0	0	0	0	1
Elections & Taxes	754	0.27	0.38	0	0	0	0.52	1

APPENDIX III

TRAINING SAMPLE (4-FOLT TIME SERIES CROSS-VALIDATION) RESULTS

Feature Set	Best Model	Accuracy	Precision	Recall	F1 Score	AUROC	Strategy PnL	Sharpe Ratio
vadertrump_lags	RF (depth=8)	0.557203	0.598616	0.650376	0.623423	0.532721	0.541378	2.090464
financial_vadernews	RF (depth=4)	0.527542	0.672000	0.315789	0.429668	0.526863	0.384642	1.478911
everything_no_lags	RF (depth=12)	0.544492	0.604082	0.556391	0.579256	0.520768	0.344699	1.324204
financial_vadertrump_lags	KNN (k=5)	0.525424	0.586066	0.537594	0.560784	0.532429	0.311641	1.196446
financial_vadertrump	XGB (lr=0.1, depth=3)	0.510593	0.640000	0.300752	0.409207	0.512273	0.292699	1.123349
financial_trump_all	KNN (k=3)	0.514831	0.570881	0.560150	0.565465	0.500310	0.276420	1.060585
financial_trumpthemes	KNN (k=3)	0.533898	0.589844	0.567669	0.578544	0.560278	0.257364	0.987176
financial_lags	RF (depth=4)	0.519068	0.648855	0.319549	0.428212	0.534674	0.252339	0.967831
trumpthemes	KNN (k=5)	0.563559	0.623967	0.567669	0.594488	0.588218	0.216599	0.830344
financial_vadernews_lags	RF (depth=12)	0.527542	0.599078	0.488722	0.538302	0.511798	0.207038	0.793598
trump_all	Logistic (C=10)	0.525424	0.577206	0.590226	0.583643	0.520093	0.205510	0.787727
financial_trump_all_snwnews	KNN (k=3)	0.536017	0.593625	0.560150	0.576402	0.529144	0.190671	0.730725
trumpthemes_lags	KNN (k=5)	0.510593	0.560554	0.609023	0.583784	0.492846	0.180219	0.690589
vadertrump	RF (depth=4)	0.514831	0.564460	0.609023	0.585895	0.517629	0.162606	0.622986
financial_trumpthemes_lags	KNN (k=3)	0.525424	0.580769	0.567669	0.574144	0.526480	0.145763	0.558373
financial_trump_all_vadernews	KNN (k=5)	0.506356	0.561798	0.563910	0.562852	0.517182	0.145574	0.557646
vadernews_lags	KNN (k=3)	0.548729	0.588040	0.665414	0.624339	0.533670	0.141685	0.542734
financial	XGB (lr=0.05, depth=4)	0.497881	0.607407	0.308271	0.408978	0.502747	0.120207	0.460385
financial_swnnews	KNN (k=3)	0.510593	0.575107	0.503759	0.537074	0.503796	0.111436	0.426766
swnnews	XGB (lr=0.1, depth=3)	0.521186	0.567114	0.635338	0.599291	0.491258	0.066897	0.256136
vadernews	Logistic (C=0.1)	0.514831	0.570342	0.563910	0.567108	0.535222	0.023606	0.090373
everything	RF (depth=12)	0.506356	0.558304	0.593985	0.575592	0.486240	0.004104	0.015712
financial_swnnews_lags	RF (depth=4)	0.493644	0.604651	0.293233	0.394937	0.511552	-0.018157	-0.069513
swnnews_lags	RF (depth=12)	0.504237	0.553333	0.624060	0.586572	0.472507	-0.045864	-0.175594

APPENDIX IV

TEST SAMPLE RESULTS

Feature Set	Best Model	Accuracy	Precision	Recall	F1 Score	AUROC	Strategy PnL	Sharpe Ratio
vadertrump	RF (depth=12)	0.61	0.69	0.63	0.66	0.61	27.4%	3.85
financial_trumpthemes	KNN (k=5)	0.55	0.67	0.48	0.56	0.56	13.6%	1.87
vadernews_lags	KNN (k=5)	0.55	0.62	0.62	0.62	0.57	10.6%	1.45
swnnews_lags	RF (depth=12)	0.57	0.64	0.68	0.66	0.51	10.0%	1.37
financial_swnnews	KNN (k=5)	0.54	0.64	0.53	0.58	0.52	5.9%	0.81
financial_vadertrump_	KNN (k=5)	0.55	0.78	0.36	0.49	0.60	5.5%	0.75
everything	RF (depth=12)	0.52	0.74	0.31	0.44	0.55	5.4%	0.74
trump_all	Logistic (C=0.1)	0.50	0.62	0.44	0.52	0.51	3.9%	0.54
trumpthemes_lags	KNN (k=5)	0.53	0.60	0.62	0.61	0.49	2.8%	0.38
vadertrump_lags	RF (depth=12)	0.55	0.64	0.54	0.59	0.54	1.2%	0.16
financial_trump_all_snwnews	KNN (k=5)	0.51	0.62	0.50	0.55	0.51	-0.5%	-0.07
vadernews	Logistic (C=0.1)	0.47	0.55	0.57	0.56	0.39	-2.2%	-0.29
everything_no_lags	RF (depth=12)	0.45	0.65	0.17	0.27	0.50	-3.5%	-0.48
financial_trump_all_vadernews	KNN (k=5)	0.50	0.65	0.36	0.46	0.51	-4.6%	-0.63
swnnews	XGB (lr=0.05, depth=4)	0.52	0.60	0.59	0.60	0.49	-5.1%	-0.69
financial_vadertrump	XGB (lr=0.05, depth=4)	0.42	1.00	0.03	0.06	0.61	-6.9%	-0.94
financial_trumpthemes	KNN (k=5)	0.45	0.55	0.41	0.47	0.43	-7.2%	-0.98
trumpthemes	KNN (k=5)	0.48	0.58	0.49	0.53	0.51	-11.7%	-1.60
financial_lags	RF (depth=12)	0.41	1.00	0.02	0.04	0.54	-12.3%	-1.69
financial	XGB (lr=0.05, depth=4)	0.41	0.60	0.03	0.06	0.61	-12.6%	-1.73
financial_vadernews	RF (depth=12)	0.40	0.50	0.01	0.02	0.47	-13.0%	-1.79
financial_trump_all	KNN (k=5)	0.45	0.56	0.41	0.47	0.46	-13.5%	-1.86
financial_swnnews_lags	RF (depth=12)	0.40	0.00	0.00	0.00	0.62	-15.1%	-2.08
financial_vadernews_lags	RF (depth=12)	0.40	0.50	0.01	0.02	0.49	-16.1%	-2.21