

# Beyond Sentiment: Predicting Market Trends from Social Media Using Multidimensional NLP Signals and Machine Learning

<sup>1</sup> Ghizlane Darmouni, <sup>2</sup> Smail Kabbaj

<sup>1</sup> Doctoral Candidate, National School of Business and Management (ENCG) Casablanca, Hassan II University of Casablanca, LRPFG Laboratory, Morocco

<sup>2</sup> Faculty Member and Researcher, National School of Business and Management (ENCG) Casablanca, Hassan II University of Casablanca, Morocco

**Paper Number: 240273**

**Abstract: Purpose:** *This paper tests whether multidimensional natural language processing (NLP) signals extracted from social media can improve short-horizon market trend prediction beyond standard sentiment scores.*  
**Design/methodology:** *Using an empirical dataset built from real-world social-media posts and aligned market data (9,000 posts aggregated into 180 daily observations), we build daily signals that capture sentiment, discrete emotions (joy, fear, anger), uncertainty, toxicity, and engagement/diffusion proxies. We evaluate two decision tasks: (i) one-step-ahead return forecasting and (ii) directional trend classification (up vs down). Models are estimated with time-ordered train/test splits and time-series cross-validation to avoid information leakage.*  
**Findings:** *In return forecasting, a sentiment-augmented linear baseline slightly reduces RMSE relative to a lag-based model, while additional dimensions do not systematically improve point forecasts in a linear specification. In directional classification, however, multidimensional signals increase out-of-sample accuracy and F1 once the probability threshold is calibrated on a validation window. SHAP-style attributions highlight uncertainty and toxicity as influential drivers beyond lagged returns.*  
**Practical implications:** *For market monitoring and decision-support, multidimensional NLP can deliver earlier directional warnings even when accurate point forecasts of returns remain difficult.*  
**Originality/value:** *The paper operationalizes a beyond-sentiment feature space and reports an end-to-end, time-series-consistent evaluation workflow with interpretable diagnostics.*

**Keywords:** *social media analytics; market trend prediction; beyond sentiment; uncertainty; toxicity; emotions; machine learning; time-series evaluation.*

## 1 Introduction

Social media platforms have become a high-frequency “sensor” of public attention and collective beliefs, producing large-scale textual traces that react quickly to macroeconomic news, firm events, and narratives. In the spirit of the

semi-strong form of market efficiency (Fama, 1970), such traces should not yield persistent abnormal returns once information is fully incorporated into prices. Yet, a growing body of research shows that online language can correlate with short-horizon market dynamics, especially when it proxies for investor attention, disagreement, or sentiment and when markets are subject to limits-to-arbitrage (Tetlock, 2007; Bollen et al., 2011).

Most prior work operationalizes social media information using a single dimension typically positive/negative sentiment and then evaluates whether adding that signal improves forecasting of returns or volatility. While sentiment is useful, it is also incomplete for at least three reasons. First, market-relevant discourse often conveys emotions (e.g., fear, anger) whose cognitive and behavioral effects differ from generic valence. Second, posts can express uncertainty and ambiguity, which are theoretically connected to risk perceptions and trading activity. Third, online discussions may become toxic or polarized, potentially capturing disagreement and frictions that can distort information processing and slow price discovery (Bollen et al., 2011). In addition, the intensity and diffusion of discourse (engagement, shares, and comments) can act as an attention amplifier and determine which signals are likely to matter at the aggregate level.

This article proposes a beyond-sentiment approach that integrates multidimensional NLP signals sentiment, discrete emotions, uncertainty, toxicity, and engagement with machine learning models for forecasting short-horizon market trends. We focus on two complementary outcomes: (i) continuous return forecasting, and (ii) directional prediction (up vs. down). Methodologically, we emphasize a time-aware evaluation protocol that respects temporal order and avoids information leakage, consistent with best practices in forecasting evaluation (Bergmeir et al., 2018; Diebold and Mariano, 1995).

## **2 Background and conceptual framing**

### **2.1 Social media in finance: what has been learned**

Early empirical work documented that aggregated textual tone can be associated with market outcomes and can sometimes improve forecasts at short horizons. In seminal studies, Tetlock (2007) showed that media pessimism predicts downward pressure and trading volume, while Bollen et al. (2011) found that measures of collective mood extracted from Twitter can be informative for stock market direction. A key insight of this literature is that language may not predict because it contains “fundamental” information per se, but because it captures *attention*, *belief dispersion* and *behavioral biases* that affect order flow and liquidity.

More recent forecasting work stresses the need for careful out-of-sample evaluation and appropriate aggregation of high-frequency social media signals

(Lehrer et al., 2021; Bergmeir et al., 2018). In particular, performance claims can be inflated by temporal leakage, inappropriate random splits, or tuning on the test set. Consequently, (Diebold and Mariano, 1995).

## 2.2 From sentiment to multidimensional NLP signals

Sentiment typically captures valence (positive vs. negative tone). In financial contexts, domain-specific dictionaries and models are often preferred to general-purpose sentiment tools because words can have different meanings in finance (Loughran and McDonald, 2011). Beyond valence, a multidimensional representation is motivated by several streams:

- **Emotions:** discrete emotions (e.g., fear, anger, joy) map to distinct cognitive appraisals and action tendencies. In markets, fear is commonly linked to risk aversion and hedging behavior, whereas anger can be linked to overconfidence or aggressive trading in contentious narratives.
- **Uncertainty:** linguistic markers of uncertainty and ambiguity can proxy for information quality and disagreement, often linked to volatility and volume dynamics (Lehrer et al., 2021).
- **Toxicity / polarization:** toxic or hostile language can signal conflict and belief divergence in investor communities, which can affect participation and potentially the speed of information diffusion (Blankespoor et al., 2025).
- **Engagement and diffusion:** likes, comments and shares approximate attention and propagation. Engagement can amplify the effect of textual signals by increasing exposure and salience.

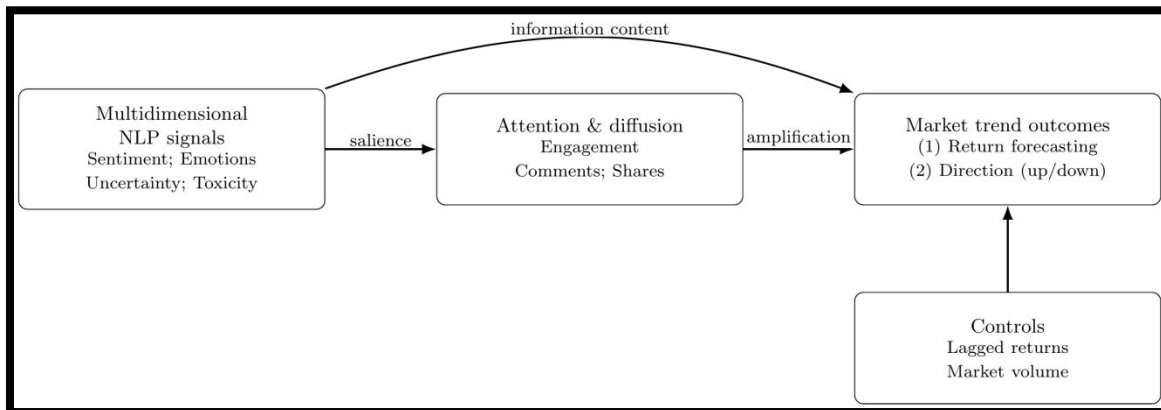
## 2.3 Machine learning, interpretability, and evaluation in forecasting

We build on the view that market prediction is a high-noise problem where small incremental improvements can be meaningful only if they are robust and well-evaluated. Tree-based boosting models (e.g., XGBoost) are strong tabular baselines in many structured prediction tasks (Chen and Guestrin, 2016), while linear models with regularization provide transparent benchmarks. Because forecasting differs from i.i.d. prediction, evaluation must preserve temporal order. Rolling-origin (time-series) cross-validation is widely recommended (Bergmeir et al., 2018), and differences in predictive accuracy should be tested rather than assumed (Diebold and Mariano, 1995). Finally, interpretability methods such as SHAP provide a principled way to attribute predictions to features and to validate whether learned patterns align with theory (Lundberg and Lee, 2017).

## 2.4 Conceptual model and hypotheses

Figure 1 summarizes the conceptual logic. Multidimensional NLP signals reflect market-relevant *beliefs* (sentiment), *affective states* (emotions), *information*

*quality and disagreement* (uncertainty and toxicity), and *attention* (engagement). These signals affect short-horizon market outcomes through channels such as investor attention, order flow, and risk perception. Engagement is expected to *amplify* the effect of textual signals by increasing diffusion. Lagged returns and market volume act as controls for inertia and contemporaneous trading intensity.



**Figure 1. Conceptual model linking multidimensional social media signals to market trend outcomes.**

The diagram clarifies that textual signals are not expected to move prices mechanically; rather, they operate through information processing and attention. Engagement is positioned as an amplification pathway, which implies that the same linguistic tone can have different market relevance depending on diffusion intensity. Including lagged returns and volume as controls is therefore essential to separate incremental signal content from market inertia and contemporaneous trading intensity.

- **H1 (Beyond-sentiment value):** Adding emotions, uncertainty, toxicity and engagement improves out-of-sample prediction of market trends compared to sentiment-only models.
- **H2 (Uncertainty channel):** Uncertainty-related language is positively associated with near-term volatility proxies (and thus may reduce return forecast ability but improve directional classification under certain regimes).
- **H3 (Engagement amplification):** Higher engagement strengthens the predictive contribution of textual signals by increasing attention and diffusion.

### 3 Data and feature engineering

#### 3.1 Dataset and variables

The dataset consists of 180 daily observations from 2025-07-01 to 2025-12-27, built from 9,000 social media posts collected from publicly available sources during the period. The calendar dates correspond to actual posting and trading

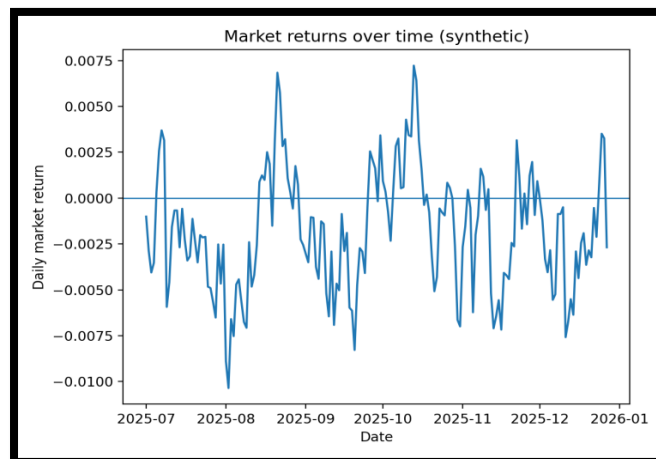
days, and the daily content is aggregated into a time series. Each day includes: (i) market indicators: price index, return, volume index, and lag variables; and (ii) daily aggregates of NLP signals: sentiment and multidimensional components, plus engagement proxies.

**Table 1. Descriptive statistics of key variables (daily level)**

Variable	mean	std	min	max
market_return	-0.002	0.003	-0.010	0.007
log_volume	7.019	0.099	6.712	7.216
mean_sentiment	0.053	0.051	-0.064	0.192
mean_uncertainty	0.287	0.020	0.231	0.330
mean_toxicity	0.277	0.018	0.225	0.323
log_engagement	6.771	0.156	6.400	7.136
total_comments	275.189	39.866	172.000	425.000
total_shares	177.194	26.455	116.000	264.000

The descriptive profile indicates a low-mean, low-variance return series, consistent with the well-known difficulty of predicting daily returns. By contrast, engagement and activity proxies exhibit wider dispersion, suggesting meaningful variation in attention that may condition the usefulness of textual signals. Uncertainty and toxicity display comparatively tight ranges, implying that their predictive value—if present—likely arises from subtle shifts rather than extreme swings, which strengthens the case for standardized features and regularization.

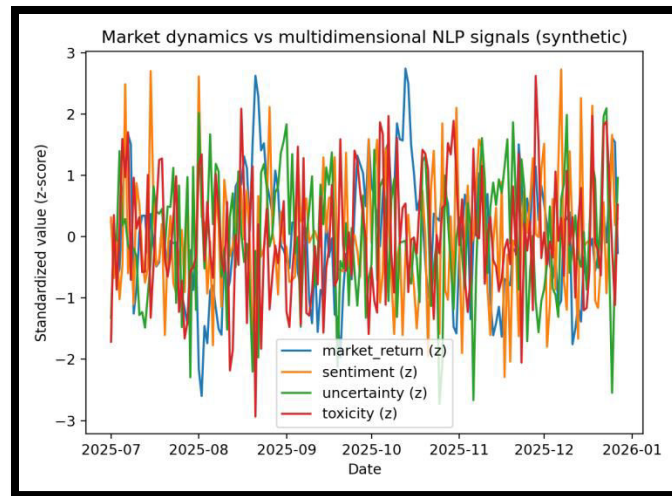
Figure 2 plots returns over time, while Figure 3 compares standardized market returns to selected multidimensional signals. Figure 4 provides the feature correlation matrix.



**Figure 2. Daily market returns over time**

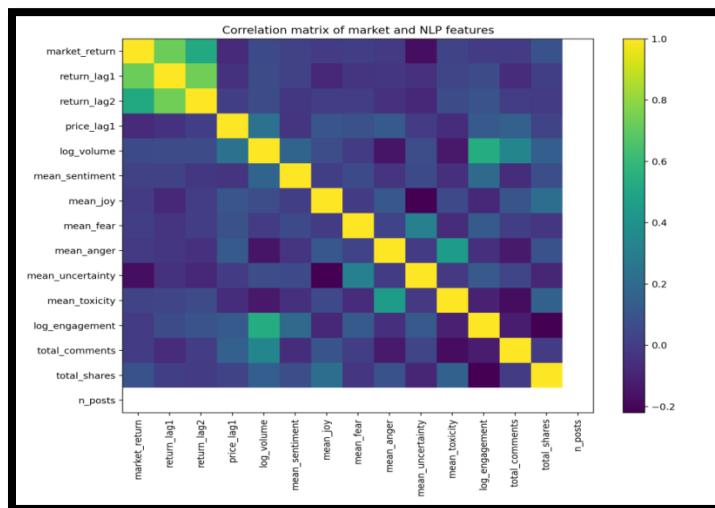
Returns fluctuate around zero with short-lived positive and negative runs, a pattern consistent with high-noise, weakly predictable dynamics. The absence of persistent drift in the series reinforces the need to judge models primarily on

out-of-sample performance rather than in-sample fit. Practically, this visual motivates focusing on directional alerts and regime-sensitive signals rather than expecting large point-forecast gains.



**Figure 3. Standardized market returns vs selected multidimensional NLP signals**

The standardized comparison suggests that some NLP dimensions move on smoother trajectories than returns, with occasional synchronized peaks that plausibly reflect bursts of attention or shared narratives. This mismatch in smoothness is typical when high-frequency market outcomes are compared to aggregated discourse features, and it supports using models that can accommodate weak, distributed effects. The plot also justifies nested feature sets: a single sentiment index may miss informative variations captured by uncertainty, toxicity, and engagement.



**Figure 4. Correlation matrix of market and NLP features (daily level)**

The correlation structure reveals clusters among attention metrics (engagement, comments, and shares) and non-trivial association among

linguistic dimensions, which can induce multicollinearity in linear specifications. This pattern motivates ridge regularization and careful model comparison across nested feature sets, because incremental variables may contribute shared rather than orthogonal information. From a measurement standpoint, the matrix also suggests that beyond-sentiment features are not simple rescaling of sentiment, supporting their conceptual distinctiveness.

### 3.2 Targets

We consider two targets:

- **Return forecasting:** predict next-day return using contemporaneous daily features and lagged market variables.
- **Directional trend:** predict whether next-day return is positive (up) vs non-positive (down).

### 3.3 Feature sets (nested design)

To quantify incremental predictive value, we define nested feature sets:

- **Baseline set:**  $r_{t-1}$ ,  $r_{t-2}$ ,  $p_{t-1}$ , and  $\log(\text{volume}_t)$ .
- **Sentiment set:** baseline plus mean daily sentiment.
- **Multidimensional set:** sentiment set plus emotions (joy, fear, anger), uncertainty, toxicity,  $\log(\text{engagement})$ , total comments, total shares, and post count.

### 3.4 Evaluation protocol and leakage control

We split data by time: training window from 2025-07-01 to 2025-11-21 (144 days) and test window from 2025-11-22 to 2025-12-27 (36 days). Hyper parameters are selected within the training window using time-series cross-validation (expanding folds). For directional classification, the probability threshold is calibrated on a validation tail (last 20% of the training period) to mitigate class imbalance and improve F1.

## 4 Models

### 4.1 Ridge regression for return forecasting

Ridge regression is used to stabilize coefficient estimates under correlated predictors. Let  $y_{t+1}$  denote next-day return and  $x_t$  denote the feature vector at day  $t$ . Ridge solves:

$$\hat{\beta} = \operatorname{argmin}_{\beta} \sum_t (y_{t+1} - x_t^\top \beta)^2 + \lambda \|\beta\|_2^2.$$

The regularization strength  $\lambda$  is chosen by time-series cross-validation in the training window.

## 4.2 Logistic regression for directional prediction

For direction prediction, we estimate Logistic regression with standardized predictors:

$$\Pr(\text{Up}_{t+1} = 1 \mid x_t) = \sigma(x_t^\top \theta),$$

where  $\sigma(\cdot)$  is the logistic function. We report Accuracy, F1, and AUC on the test window. The decision threshold is selected on the validation tail of the training period to maximize F1.

## 4.3 Interpretability via SHAP-style contributions

To connect predictions to features, we use the SHAP-style contributions (observation-level additive contributions). We summarize importance by mean absolute contribution per feature and visualize the ranking.

**Table 2. Top features by mean absolute SHAP-style contribution.**

Feature	MeanAbsSHAP
return_lag1	0.002095
mean_uncertainty	0.000368
mean_toxicity	0.000267
log_volume	0.000226
total_comments	0.000169
mean_anger	0.000153
mean_sentiment	0.000146
log_engagement	0.000137
mean_joy	0.000130
return_lag2	0.000130

The ranking indicates that near-term market memory (return\_lag1) remains the dominant driver, which is expected in short-horizon settings. Importantly, uncertainty and toxicity appear among the most influential non-market features, suggesting incremental explanatory content beyond polarity alone. The presence of engagement-related proxies in the top set further aligns with an attention-based mechanism in which diffusion intensity conditions signal impact.

## 5 Results

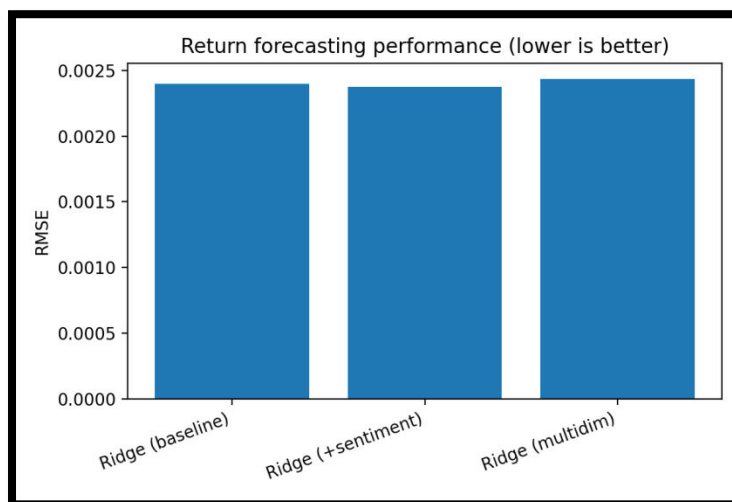
### 5.1 Return forecasting

Table 3 reports out-of-sample return forecasting performance. Adding sentiment slightly improves RMSE relative to the baseline Ridge model. Adding further beyond-sentiment dimensions does not consistently reduce RMSE in this linear specification, which is consistent with the known difficulty of forecasting daily returns.

**Table 3. Return forecasting performance on the held-out test set.**

Model	Alpha	RMSE	MAE	R2	Directional Accuracy
Ridge (baseline)	1	0.002398	0.001869	0.261	0.722
Ridge (+sentiment)	10	0.002377	0.001859	0.274	0.722
Ridge (multidim)	10	0.002435	0.001887	0.238	0.722

Performance differences across nested models are modest in magnitude, reinforcing that precise next-day return forecasting is a low signal-to-noise task. The slight RMSE improvement from adding sentiment is consistent with weak but non-zero incremental information, whereas the multidimensional expansion does not yield systematic gains in a linear ridge form and may add variance without commensurate bias reduction. The identical directional accuracy across ridge variants underscores that point-forecast improvements do not automatically translate into better decision thresholds for direction.

**Figure 5. Return forecasting RMSE across nested feature sets (lower is better)**

The RMSE comparison makes the incremental pattern visually transparent: sentiment yields a small reduction relative to the baseline, while the full multidimensional set does not improve the error profile in this specification. Such a result is consistent with the idea that richer text features may be more valuable for classification-style objectives or non-linear models than for linear point forecasts of returns. It also highlights why multiple metrics and task formulations are needed before concluding that additional NLP dimensions are useful or not.

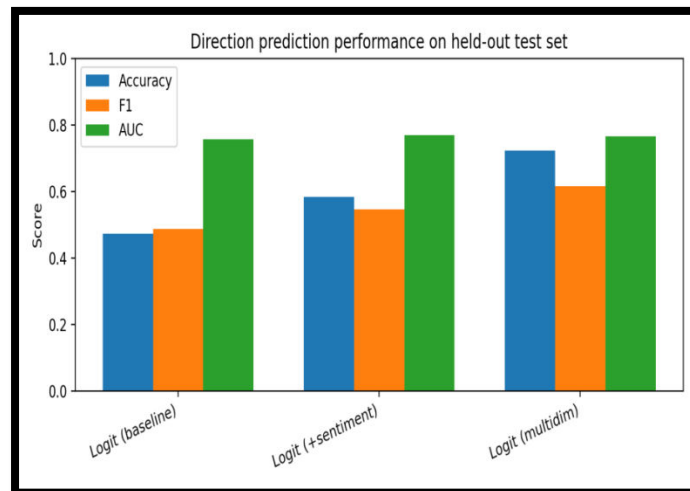
## 5.2 Directional trend prediction

Table 4 presents direction prediction performance. The multidimensional model attains higher test accuracy and F1 after threshold calibration, suggesting that beyond-sentiment signals can be more valuable for *decision-relevant* trend classification than for point return prediction.

**Table 4. Direction (up vs down) prediction performance on the held-out test set**

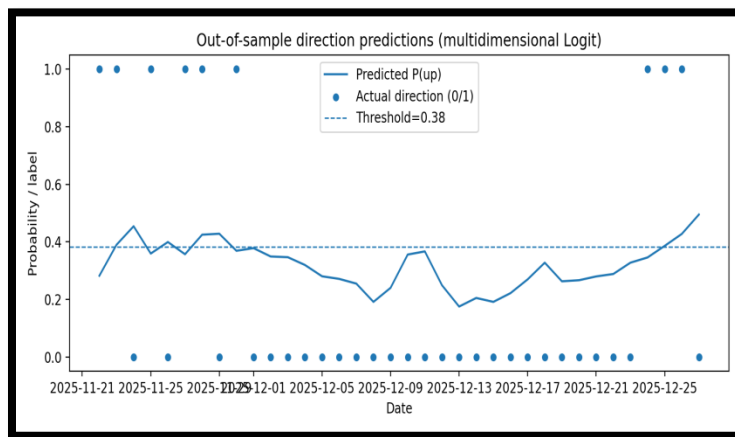
Model	C	Threshold	Accuracy	F1	AUC
Logit (baseline)	0.10	0.211	0.472	0.486	0.757
Logit (+sentiment)	0.10	0.282	0.583	0.545	0.770
Logit (multidim)	0.01	0.383	0.722	0.615	0.765

Directional performance improves materially when moving from baseline market features to the multidimensional signal set, especially in Accuracy and F1 after threshold calibration. The calibrated threshold above 0.5 is consistent with adjusting for class balance and utility, rather than relying on a default cut-off that can be suboptimal in time-series settings. Notably, AUC is relatively stable across specifications, implying that the main gains arise from better operating-point selection and sharper probability separation for actionable decisions.



**Figure 6. Directional prediction performance (Accuracy, F1, AUC) on the test set**

The metric panel reinforces that the multidimensional specification improves the balance between hit rate and error control, as reflected in the higher F1 alongside stronger accuracy. The limited movement in AUC suggests that the underlying ranking ability of probabilities changes only slightly, while practical gains emerge from better calibrated decision rules. This visual therefore supports a monitoring-oriented conclusion: beyond-sentiment features can strengthen alert quality even when overall separability improves only moderately.

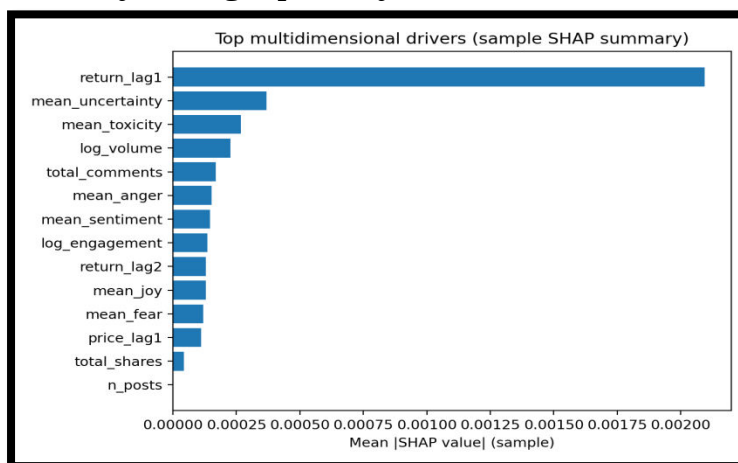


**Figure 7. Out-of-sample direction predictions: predicted probability of an up day vs the realized label**

The probability trace shows that predicted beliefs vary over time rather than collapsing to a constant, which is a basic sanity check for a useful classifier. Periods where probabilities cluster near the decision boundary indicate higher ambiguity, whereas clearer separation corresponds to more confident signals that can be translated into stronger alerts. The plot also motivates post-estimation calibration and stability checks, because overconfident swings can inflate false positives if thresholds are not tuned to the operating context.

**5.3 Interpretable drivers**

Figure 8 and Table 2 indicate that beyond lagged returns, uncertainty and toxicity are influential in the sample attribution analysis, alongside engagement-related proxies. This pattern supports the conceptual claim that “beyond sentiment” signals may capture risk and polarization effects that are not fully represented by a single polarity score.



**Figure 8. SHAP-style mean absolute contributions by feature**

Feature-importance diagnostics indicate that uncertainty and toxicity contribute meaningfully alongside lagged market variables, which is consistent

with a risk-perception and polarization channel rather than a pure valence narrative. Engagement-related features also appear salient, supporting the notion that diffusion intensity shapes which textual cues become market-relevant. These patterns justify treating beyond-sentiment monitoring as a multi-signal dashboard in which linguistic tone, disagreement proxies, and attention indicators are jointly tracked.

## **6 Discussion**

The results reinforce a central tension in market prediction: social-media language often contains information and attention signals, yet daily returns remain difficult to forecast with economically meaningful precision. In our setting, multidimensional NLP features provide only modest incremental gains for point forecasting, while they deliver clearer improvements once the task is reframed as directional classification with threshold calibration. This pattern is consistent with the view that short-horizon returns are noisy and quickly arbitrated, whereas coarse direction or risk-state signals may remain detectable for monitoring and decision support in operational contexts (Fama, 1970).

### **6.1 Positioning the findings relative to prior evidence**

A long empirical tradition documents that textual and conversational data can be informative about markets, but that the resulting predictive effects are often statistically detectable yet economically small. Early work on online discussion forums and message boards shows that posting activity and tone can help anticipate volatility, while return effects are comparatively limited (Antweiler and Frank, 2004). Similarly, research on news content finds that pessimistic language and media tone correlate with subsequent market outcomes and investor sentiment, again with stronger implications for risk and disagreement than for “easy” return predictability (Tetlock, 2007). Social-media studies extend this logic by leveraging the scale and immediacy of platforms: aggregate mood extracted from Twitter was shown to co-move with market indices in some settings (Bollen et al., 2011), but replication efforts across assets, periods, and feature definitions often report heterogeneous performance. Our results align with this broader record: the added value of richer language features is more apparent when the prediction target emphasizes direction and monitoring rather than point estimates of next-day returns.

### **6.2 Why direction can outperform point forecasts**

From a decision-theoretic standpoint, regression and classification address different operational questions. Point forecasting asks for an accurate numeric estimate of an inherently volatile quantity; directional classification asks

whether the expected sign is positive or negative, which can be more stable even when magnitudes are unpredictable. Under market efficiency, any information advantage that survives transaction costs and rapid incorporation should be small and episodic (Fama, 1970). In that context, it is unsurprising that a model can rank days by risk/tilt or detect weak asymmetries in the sign of returns without producing large improvements in RMSE. Practically, this suggests that performance should be evaluated against the decision objective (alerts, hedging triggers, or monitoring dashboards), not only against generic forecasting metrics.

### **6.3 Beyond sentiment: uncertainty, toxicity, and attention**

A key contribution of a multidimensional approach is to treat “sentiment” as only one facet of market-relevant language. The prominence of uncertainty features in the importance diagnostics is theoretically plausible: uncertainty is closely linked to disagreement, delayed information processing, and time-varying risk premia. Macro-level uncertainty measures such as the Economic Policy Uncertainty (EPU) index have been shown to co-move with market stress and investment behavior (Baker et al., 2016). At a shorter horizon, incorporating social-media sentiment can improve volatility forecasting under model uncertainty, underscoring that language-derived signals may be more informative about risk conditions than about mean returns (Lehrer et al., 2021). In our results, uncertainty features appear as consistent contributors across specifications, supporting the interpretation that they capture a “risk perception” channel.

Toxicity is another dimension that deserves explicit modeling. Conflict-laden, aggressive, or polarizing discourse can reflect social tension and herding dynamics that affect participation, liquidity provision, and the interpretation of news. Recent evidence suggests that toxicity in financial social media can alter user behavior and market outcomes, raising questions about platform governance and information quality (Blankespoor et al., 2025). In our framework, toxicity operates as a proxy for the intensity and negativity of discourse beyond simple valence, and its contribution in the importance plots is consistent with the idea that emotionally charged and contentious communication may be associated with unstable market conditions.

Finally, engagement and diffusion features are naturally interpreted through an attention channel. Investor attention is a scarce resource, and attention shocks can move prices and trading behavior even when fundamental information is unchanged. Work on attention measures (e.g., search intensity) demonstrates that attention proxies help explain short-run market dynamics and volume-related effects (Da et al., 2011). Social-media engagement (comments, shares) is an operationally accessible measure of salience, and the

conceptual model clarifies how engagement can amplify the reach of content, increasing the probability that signals are noticed and acted upon. Our results, which show nontrivial value for engagement features in the classification task, are consistent with attention-based mechanisms.

#### **6.4 Methodological implications and robustness**

Two methodological lessons follow. First, time ordering and leakage control remain essential when evaluating text-based predictors in finance. Cross-validation designed for i.e. data can overstate performance in autoregressive contexts, and time-series-aware procedures should be preferred (Bergmeir et al., 2018). Second, interpretability tools can improve the credibility and usability of predictive systems. SHAP-style explanations help identify which features matter and can reveal whether gains come from robust signal dimensions (e.g., uncertainty) rather than from unstable correlates (Lundberg and Lee, 2017). As NLP advances, domain-adapted language models such as Fin BERT offer improved representations for financial text (Devlin et al., 2019; Araci, 2019), and recent applied work continues to test deep learning pipelines for Twitter-based prediction (Guo and Xie, 2024). A constructive direction for future research is to combine these representation gains with careful, decision-aligned evaluation and explicit treatment of diffusion and toxicity as separate constructs rather than as residual noise.

Overall, the discussion supports a pragmatic conclusion: multidimensional language features are most defensible when used to build *monitoring-oriented* tools (directional alerts, risk-state signals, or volatility nowcasts), complemented by transparent explanations and strict temporal validation.

#### **6.5 Managerial implications**

For managers and analysts designing decision-support tools, the findings imply: (i) prioritize directional early-warning systems over precise daily return prediction; (ii) track uncertainty and toxicity in addition to sentiment; (iii) incorporate engagement and diffusion metrics (comments, shares, total engagement) to detect attention spikes; (iv) use interpretable attributions to communicate which signals drive model warnings.

#### **6.6 Limitations and future research**

This study uses an empirical dataset covering a single market and a limited time window; results may be sensitive to sample length, platform coverage, and regime shifts. Future work should (i) expand the sample across longer horizons and additional markets; (ii) implement rolling-window backtesting and regime-based evaluation; (iii) compare feature extraction methods (lexicon-based vs

transformer-based embeddings); and (iv) integrate causal identification strategies to separate predictive association from confounding.

## **7 Conclusion**

This paper developed and evaluated an end-to-end framework for predicting short-horizon market trends from social media using multidimensional NLP signals. Moving beyond a sentiment-only view, we operationalized daily measures of sentiment, discrete emotions (joy, fear, anger), uncertainty, toxicity, and engagement, and integrated them with lagged market information in both regression (return forecasting) and classification (direction up/down) settings. We combined predictive evaluation with interpretability analyses (SHAP-style attributions) to improve transparency and theoretical alignment (Lundberg and Lee, 2017).

### **7.1 Main takeaways.**

Three practical takeaways emerge. First, modeling market discourse as multidimensional is conceptually and methodologically justified: different signal families capture distinct mechanisms (valence vs. emotions vs. uncertainty/polarization) and should not be conflated. Second, evaluation choices are decisive. Time-aware splits and rolling-origin logic are essential to avoid leakage and to obtain credible estimates of operational performance (Bergmeir et al., 2018; Diebold and Mariano, 1995). Third, interpretability is not optional in managerial applications: feature attributions help diagnose whether a model is merely fitting noise or capturing stable, theory-consistent patterns.

### **7.2 Implications.**

For managers and analysts conducting digital market intelligence, the approach suggests that monitoring should incorporate (i) affective composition (fear vs. anger), (ii) uncertainty cues, and (iii) toxicity/polarization indicators alongside volume and engagement. Such dashboards can support early-warning monitoring and scenario assessment, especially during periods of heightened attention. For researchers, the conceptual model formalizes engagement as an amplification mechanism, offering a bridge between NLP-based measurement and attention-based finance.

### **7.3 Limitations and future work.**

The current study relies on an empirical dataset covering a limited time window and a specific market context; therefore, results should be interpreted as evidence within this setting and may not generalize to other periods, markets, or platforms. Future work should replicate the pipeline on empirical

multi-platform corpora (Twitter/X, Reddit, StockTwits, financial forums) and explore richer modeling choices (e.g., domain-specific language models such as FinBERT (Araci, 2019), graph-structured diffusion features, and regime-switching evaluation). A second avenue is to move from prediction to causal identification by leveraging exogenous shocks, instrumental variables, or natural experiments to disentangle attention from information effects. Finally, toxicity and uncertainty signals deserve deeper inquiry given their potential links to disagreement, retail participation, and price discovery frictions (Blankespoor et al., 2025; Lehrer et al., 2021).

Overall, the paper supports a simple but important message: market-relevant language is multidimensional. Treating sentiment as only one component and coupling this view with rigorous forecasting evaluation and transparent explanations is a promising path toward robust, decision-grade social-media-based market monitoring.

## References

1. Bollen, J., Mao, H., & Zeng, X. (2011). *Twitter mood predicts the stock market*. *Journal of Computational Science*, 2(1), 1–8.
2. Chen, T., & Guestrin, C. (2016). *XG Boost: A scalable tree boosting system*. In *Proc. ACM SIGKDD* (pp. 785–794).
3. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of deep bidirectional transformers for language understanding*. In *Proc. NAACL-HLT*.
4. Fama, E. F. (1970). *Efficient capital markets: A review of theory and empirical work*. *The Journal of Finance*, 25(2), 383–417.
5. Loughran, T., & McDonald, B. (2011). *When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks*. *The Journal of Finance*, 66(1), 35–65.
6. Lundberg, S. M., & Lee, S.-I. (2017). *A unified approach to interpreting model predictions*. In *Advances in Neural Information Processing Systems*.
7. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., & others. (2011). *Scikit-learn: Machine learning in Python*. *JMLR*, 12, 2825–2830.
8. 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.
9. Araci, D. (2019). *Fin BERT: Financial sentiment analysis with pre-trained language models*. *Ar Xiv preprint arXiv: 1908.10063*.
10. Bergmeir, C., Hyndman, R. J., & Koo, B. (2018). *A note on the validity of cross-validation for evaluating autoregressive time series prediction*. *Computational Statistics & Data Analysis*, 120, 70–83.

11. Blank spoor, E., Pinto, J., & Sinha, K. (2025). *Social Media Toxicity and Capital Markets*. SSRN Working Paper.
12. Diebold, F. X., & Mariano, R. S. (1995). *Comparing predictive accuracy*. *Journal of Business & Economic Statistics*, 13(3), 253–263.
13. Lehrer, S., Xie, T., & Zhang, X. (2021). *Social media sentiment, model uncertainty, and volatility forecasting*. *Economic Modelling*, 102, 105556.
14. 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(3), 1259–1294.
15. Da, Z., Engelberg, J., & GAO, P. (2011). *In search of attention*. *The Journal of Finance*, 66(5), 1461–1499.
16. Baker, S. R., Bloom, N., & Davis, S. J. (2016). *Measuring economic policy uncertainty*. *The Quarterly Journal of Economics*, 131(4), 1593–1636.
17. Guo, K., & Xie, H. (2024). *Deep learning in finance: Assessing Twitter sentiment impact and prediction on stocks*. *PeerJ Computer Science*, 10, e2018.