Proceedings of Machine Learning Research **318** (2026)
The 39th Canadian Conference on Artificial Intelligence
Canadian AI 2026
25-29 May 2026,Vancouver,BC

Directional Stock Prediction with Temporal Sentiment

Nabil Benjaa*, Amir Ben Khalifa, Faisal Tareque Shohan
Bessam Abdulrazak, Amine Trabelsi, Shengrui Wang
Université de Sherbrooke, Sherbrooke, QC, Canada

Abstract

Financial market forecasting is increasingly incorporating textual sentiment cues from news and financial reports alongside traditional market indicators. While prior work has frequently incorporated aggregated daily sentiment indicators, the temporal structure through which sentiment propagates into market movements remains underexplored. Sentiment influence on index price dynamics may continue beyond a single observation period, thereby limiting the ability of point-in-time sentiment measures. Moreover, predictive performance is often evaluated using regression metrics such as Mean Absolute Error and Mean Absolute Percentage Error. Although these metrics provide valuable insights into prediction error, they fall short of capturing the effectiveness required in financial settings, where accurately predicting the direction of price movements is crucial. To address these limitations, we introduce temporally sentiment features that capture the persistence and evolution of market perception over time rather than relying on the last-day sentiment. In addition, we propose a Transformer-based forecasting architecture specifically designed to model temporal dependencies between sentiment and index returns. Our approach also prioritizes directional evaluation and incorporates an asymmetric custom objective function to better address the risks associated with negative market movements. Findings indicate that, while conventional error metrics are comparable to baseline models, the integration of temporal sentiment significantly enhances overall directional prediction. Furthermore, employing an asymmetric custom objective function especially in the context of the Transformer based model improves the identification of downward trends while ensuring a more effective balance between positive and negative market fluctuations.

Keywords: Forecasting, Temporal sentiment, Finance, Time series, LSTM, Transformer, Directional prediction

1. Introduction

The stock market plays a central role in the global economic system, serving as a key mechanism for capital allocation, risk transfer, and information aggregation. Major indices such as the S&P 500 and NASDAQ reflect the performance of large segments of the econ- omy and influence investment decisions made by institutions, policymakers, and individual investors [1, 2]. As a result, understanding and forecasting index movements is of substantial economic and practical importance. In addition to traditional market variables such as prices, returns, and trading volumes, financial news and public sentiment significantly affect investor behavior and market trends [3, 4]. With recent advances in natural language processing, sentiment analysis has become an important tool for extracting signals from financial texts and incorporating them into forecasting models [5, 6].

Financial markets are inherently characterized by volatility and noise. Price variations are driven not only by observable economic indicators but also by unpredictable events such as macroeconomic announcements, geopolitical crises, natural disasters, or sudden policy changes [7]. These shocks can alter expectations rapidly and generate abrupt fluctuations in returns. As a result, market behavior often exhibits nonlinear dynamics and complex interactions across time, making accurate forecasting particularly challenging [7, 8].

Existing research demonstrates that integrating news sentiment with market data can enhance stock prediction performance [9, 10]. Predominantly, existing methodologies utilize a singular daily sentiment score derived from financial news, which is then incorporated into regression models alongside historical prices or technical indicators [11]. Although these approaches can reduce numerical prediction error, they often regard sentiment as a signal relevant only for the same day.

However, the impact of news may persist beyond a single day, as information can influence market behavior across multiple periods rather than at a singular, isolated point in time [5]. For example, a major earnings surprise, a regulatory investigation, or the announcement of a global health crisis may trigger an initial price reaction, followed by continued adjustments over subsequent trading sessions as investors reassess risk,

* Nabil.Benjaa@USherbrooke.ca

rebalance portfolios, and incorporate additional related information. In such cases, treating sentiment as a purely contemporaneous signal may underestimate its temporal propagation and delayed market effects.

Another limitation of existing approaches lies in evaluation methodology. Conventional metrics such as MAE and MAPE quantify prediction error but do not fully capture the requirements of financial decision-making, where correctly anticipating the direction of price movements is often more critical than minimizing numerical deviations [11]. While the majority of prior work reports good overall recall and precision, these results often reflect a bias toward predicting positive movements, since most models tend to favor the majority class [12, 13]. This behavior can be particularly costly in finance: predicting an upward movement when the market actually declines (a false positive) can lead to substantial losses. To address this, we reformulate the task as a classification problem and implement Transformer-based models with a custom asymmetric loss function that penalizes false positives more heavily [6, 14]. Our experiments show that this approach substantially improves the detection of negative trends while maintaining reasonable performance on upward movements. Although performance is not perfectly balanced across both directions, the model avoids extreme bias toward a single class and achieves a more reliable trade-off between positive and negative predictions, particularly during volatile market periods.

We extend a baseline framework by introducing sentiment features that capture per- sistence and evolution over time and integrate them into the same predictive pipeline for controlled comparison [5, 11]. Both LSTM- and Transformer-based models are employed to assess the effectiveness of sequential architectures in capturing these temporal patterns [8, 14]. We further emphasize directional evaluation and adapt the modeling strategy to better account for asymmetric risks associated with negative market movements. Our objective is to determine whether temporally structured sentiment, combined with advanced sequential modeling, provides more informative signals and leads to improved directional forecasts in financial prediction tasks.

The main contributions of this work are summarized as follows:

- We introduce temporally structured sentiment features that effectively capture the persistence and dynamic evolution of market perception, moving beyond the limitations of single-day aggregated sentiment signals.
- We propose a Transformer-based forecasting model designed to leverage these temporal representations and model dependencies between sentiment dynamics and stock index returns.
- We emphasize directional evaluation and incorporate an asymmetric objective formulation to enhance the prediction of downward market movements, which is particularly relevant in financial contexts.

All experiments were conducted using our proposed framework. The code is available at: `https://github.com/benjaanabil/D-Stock-Prediction-with-T-Sentiment`.

2. Related Work

Machine learning methods, espacially deep learning methods have become increasingly prominent in stock market prediction due to their ability to capture nonlinear relationships in financial time series [1, 2]. Sequential models such as Long Short-Term Memory (LSTM) networks, have demonstrated strong performance in modeling temporal dependencies [7, 8]. More recently, transformer-based models leverage attention mechanisms to capture long-range dependencies and complex interactions, offering improved representation learning for financial forecasting tasks [6, 14].

Sentiment analysis from financial news has emerged as an important exogenous signal for stock prediction. Domain-specific language models such as FinBERT enable precise extraction of sentiment scores from textual data, which can then be combined with historical prices and technical indicators to enhance forecasting performance [4, 5, 9]. Early approaches generally modeled sentiment as a single daily score, which while improving overall numerical accuracy, often resulted in poor directional prediction, particularly for negative movements. The FinBERT-LSTM framework proposed by Gu et al. [11] serves as a representative baseline, integrating daily sentiment scores with historical stock prices. Despite achieving strong overall accuracy, directional evaluation revealed significant imbalance, with models favoring upward movements and underperforming on downward trends.

Subsequent studies have explored the use of transformer-based architectures and hybrid sequential models to address these limitations. These studies generally improve overall directional performance by capturing richer temporal and contextual dependencies [6, 14]. However, most of them still report a bias toward positive trends, and negative movement detection remains challenging, which is particularly costly in financial applications due to the asymmetric impact of false positives.

In summary, prior research demonstrates the benefits of integrating sentiment and leveraging advanced sequential architectures for stock prediction. Nonetheless, a common limitation is the imbalance in directional performance, especially for negative movements. Our work addresses this gap by modeling sentiment as a temporal signal and introducing a custom asymmetric loss function in both LSTM- and Transformer-based models. This explicitly improves the detection of negative trends while maintaining balanced performance across upward and downward movements, enhancing reliability in financial decision-making.

3. Methodology

This section describes the proposed framework in detail, following the pipeline shown in Figure 1.

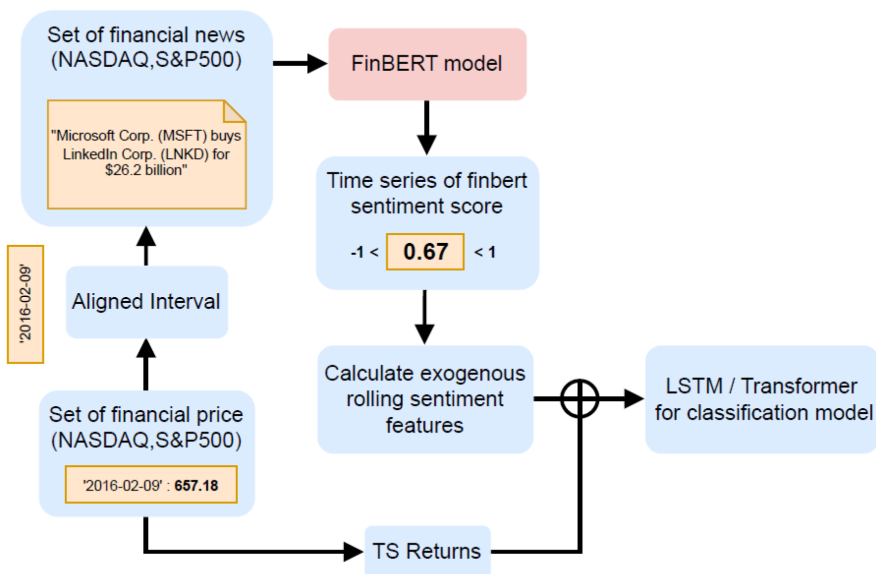

Figure 1. Overall architecture of the pipeline

Our approach builds upon the FinBERT-LSTM framework proposed by Gu et al. [11], which we use as a benchmark baseline for comparison. By enriching sentiment representation and reformulating the prediction task to better capture market direction. Our methodological approach combines a conventional forecasting framework with temporally structured sentiment features, sequential modeling techniques, and an asymmetric classification strategy.

Figure 1 presents the overall architecture of the proposed framework. The pipeline consists of four main stages. First, financial news and historical price data are collected and temporally aligned on a daily basis. Second, news articles are processed using FinBERT to generate daily sentiment scores, which are then transformed into rolling temporal sentiment features. Third, these engineered sentiment features are combined with historical price sequences to form sequential inputs. Finally, the combined sequences are fed into either an LSTM or a Transformer model to perform next-day prediction, formulated either as regression (price forecasting) or as directional classification with an asymmetric objective function.

3.1. FinBERT Score

To quantify the sentiment expressed in financial news text, we use **FinBERT**, a transformer-based language model specifically adapted to the financial domain [5]. FinBERT is built upon the BERT architecture and pre-trained on large collections of financial communications and reports, allowing it to capture domain-specific terminology and sentiment patterns that differ from general-language models. The model takes financial text sequences such as news headlines as input, which are tokenized and encoded to produce a contextual semantic representation of the entire text.

FinBERT outputs probabilities for three sentiment classes: positive, neutral, and negative. These probabilities are combined to produce a single continuous sentiment score ranging from -1 (strongly negative) to $+1$ (strongly positive), with values near zero indicating neutral sentiment. When multiple news items correspond to the same calendar date, their sentiment scores are averaged to obtain a single value aligned with the corresponding price observation. These aggregated sentiment values are then used as explanatory features in the predictive modeling framework.

3.2. **Sentiment Feature Engineering**

The baseline relies on a single sentiment value per day, which cannot capture delayed or cumulative reactions of the market to news. To address this limitation, we construct temporally-informed sentiment features derived from the daily FinBERT score S_t. Let S_t denote the sentiment score at time t, and let k be the rolling window size. The following features are computed:

Table 1. Mathematical definition of sentiment features.

Feature	Definition
Sentiment direction	$\mathrm{Si}_t = \mathrm{sign}(S_t)$
Sentiment difference	$\Delta S_t = S_t - S_{t-1}$
Momentum	$M_t^{(k)} = S_t - S_{t-k}$
Rolling mean	$\overline{S}_t^{(k)} = \frac{1}{k}\sum_{i=0}^{k-1} S_{t-i}$
Rolling volatility	$\sigma_t^{(k)} = \sqrt{\frac{1}{k}\sum_{i=0}^{k-1}(S_{t-i} - \overline{S}_t^{(k)})^2}$
Expected Sentiment Shift (ESS)	$\mathrm{ESS}_t = \frac{1}{k}\sum_{i=0}^{k-1} \mathrm{Si}_{t-i}\lvert S_{t-i}\rvert$
Sentiment Confidence Score (SCS)	$\mathrm{SCS}_t = \frac{1}{k}\sum_{i=0}^{k-1} \lvert S_{t-i}\rvert$
Weighted Sentiment Ratio (WSR)	$\mathrm{WSR}_t = \frac{\sum_{i=t-k}^{t}\max(S_i,0)}{\sum_{i=t-k}^{t}\lvert\min(S_i,0)\rvert+\epsilon}$
Sentiment Polarity Ratio (SPR)	$\mathrm{SPR}_t = \frac{\sum_{i=t-k}^{t}\max(S_i,0)-\sum_{i=t-k}^{t}\lvert\min(S_i,0)\rvert}{\sum_{i=t-k}^{t}\lvert S_i\rvert+\epsilon}$
Sentiment shock	$\mathrm{Shock}_t = \mathrm{ESS}_t - \mathrm{ESS}_{t-1}$

where:

- S_t is the FinBERT sentiment score at time t,
- Si_t represents sentiment polarity,
- k is the temporal window size,
- ϵ is a small constant preventing division by zero.

The rolling window size was set to $k = 3$ to capture the short-term impact of financial news on market behavior. Financial sentiment typically affects prices within a few trading days, making short windows more appropriate for modeling rapid reactions. From the data perspective, the sentiment series exhibits high variability and non-stationary behavior; using a small window improves the stability of derived features while limiting noise accumulation and avoiding excessive smoothing. Empirical experiments with larger windows ($k = 5$, $k = 10$) led to weaker directional performance, particularly in terms of mean directional accuracy we will discuss later. Thus, $k = 3$ provides an effective trade-off between responsiveness, stability, and noise control in temporal sentiment modeling.

To reduce dimensionality and remove redundant sentiment signals, we apply a Random Forest classifier to rank feature importance [15]. The model is trained on candidate sentiment features, and the resulting importance scores are used to select the most predictive variables. Only the five top-ranked features ‘Sentiment direction‘, ‘Sentiment difference‘, ‘Sentiment shock‘, ‘Sentiment Confidence Score‘, and ‘Expected Sentiment Shift‘ are retained and used as inputs for the subsequent LSTM or Transformer models, improving efficiency and minimizing noise.

3.3. Sequential Regression Models

Stock prediction is formulated as a sequential learning problem using time-series inputs composed of historical prices and sentiment features. Let P_t denote the daily closing price and F_t the sentiment feature vector. Each input sequence is defined as:

$$X_t = [P_{t-T+1:t},\, F_{t-T+1:t}],$$

where T is size of the look-back window.

Following the framework proposed by Gu et al. [11], which we adopt as our benchmark baseline, the initial task is formulated as a regression problem aimed at predicting the next-day closing price. In this context, it refers to the previously published FinBERT-LSTM model that combines a single daily sentiment score with historical price data. We reproduce this setup to ensure fair comparison and to evaluate whether our temporally structured sentiment features provide additional predictive value.

To show the effectiveness of temporal sentiment features, we retain the same LSTM architecture used in the baseline and extend its input space by incorporating the selected temporal sentiment features alongside historical price sequences. The model is trained to minimize prediction error between the estimated and actual next-day closing prices.

Although this regression formulation enables benchmarking against existing approaches, it is important to note that numerical accuracy alone does not ensure reliable directional forecasts. Price estimation does not necessarily translate into reliable prediction of market direction; this limitation motivates the directional modeling approach presented in the next section. For instance, a model may produce predictions that are numerically close to the true price while still failing to correctly anticipate whether the market will move upward or downward. In financial decision-making, such errors can lead to incorrect trading positions despite seemingly accurate forecasts.

3.4. Directional Classification with Asymmetric Loss

To better capture upward and downward market movements, the prediction task is reformulated from regression to binary classification, where the objective is to predict the next-day price direction rather than the exact price value. This formulation aligns more closely with financial decision-making, where correctly anticipating negative movements is particularly critical for risk management.

The baseline regression setup is reformulated as a directional classification task aimed at predicting the next-day price trend. To construct the target, daily price changes are first expressed as returns. Let P_t denote the closing price at time t. The return is computed as:

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}}.$$

The prediction target is then defined based on the sign of the next-day return:

$$y_t = \begin{cases} 1, & \text{if } r_{t+1} > 0, \\ 0, & \text{otherwise.} \end{cases}$$

This formulation enables the model to focus on directional movement rather than exact price values.

We retain the same LSTM architecture used in the regression setting and replace the final layer with a binary classification head. The model outputs the probability of an upward movement via a sigmoid activation function [16], producing values in the interval $[0, 1]$. The predicted class label is then obtained using standard probability thresholding at 0.5, where 1 denotes an upward movement and 0 denotes a downward movement. The network is trained on sequential inputs composed of historical prices and engineered sentiment features to capture temporal dependencies relevant to market direction.

3.5. Asymmetric Objective Function

Financial markets exhibit asymmetric risk, where predicting a positive movement during an actual downturn can lead to substantial financial losses. In contrast, predicting a negative movement when the market actually rises is generally more acceptable in practice, as it mainly results in a missed opportunity rather than a direct loss. This asymmetry motivates the use of an objective function that treats these two types of errors differently.

To address this issue, we introduce a custom asymmetric objective function that penalizes false positive predictions more heavily than false negatives. The loss is defined as:

$$L(y, \hat{y}) = \mathrm{BCE}(y, \hat{y}) + \lambda \cdot \hat{y}(1 - y),$$

where BCE denotes binary cross-entropy

$$\mathrm{BCE}(y, \hat{y}) = -\left[y \log(\hat{y}) + (1 - y)\log(1 - \hat{y})\right].$$

and $y \in \{0, 1\}$ represents the true direction, $\hat{y}$ is the predicted probability, and $\lambda > 1$ controls the penalty assigned to false positives. This formulation reflects financial decision-making constraints, improving sensitivity to negative market movements while allowing the model to tolerate conservative predictions that may miss some positive trends but avoid costly incorrect bullish signals.

We hypothesize that introducing an asymmetric objective within the LSTM framework will improve the detection of negative trends, which are often underrepresented in conventional models. This hypothesis is motivated by financial decision-making considerations, where failing to anticipate a market downturn can lead to direct financial losses, whereas predicting a decline during an actual upward movement mainly results in missed investment opportunities. However, we also expect that stronger penalization of false positives may reduce performance on positive movements, as the model may become more conservative and less likely to predict upward trends. The empirical validity of this hypothesis is evaluated in the experimental section.

3.6. Transformer-Based Directional Modeling

LSTM models process data sequentially, older signals fade, causing the model to default to safe, average predictions to minimize errors. The Transformer encoder avoids this limitation. By using attention mechanisms, it views the entire history of price and sentiment data at once without compressing it. This direct access allows the Transformer to confidently spot the distant patterns that trigger positive or negative movements, making accurate predictions rather than defaulting to the mean, even under the exact same loss penalty.

This configuration leads to more balanced directional performance, improving the detection of negative trends while maintaining competitive accuracy for positive movements. The combination of temporally structured sentiment, attention-based modeling, and asymmetric learning provides a more robust framework for directional financial forecasting.

3.7. Implementation

Two deep learning architectures were implemented to predict next-day market direction from historical returns and news-based sentiment signals. Daily FinBERT sentiment scores were aggregated into temporal features, and the most informative variables were selected using a Random Forest importance ranking. Input sequences of 10 consecutive days, composed of past returns and selected sentiment indicators, were used for both models.

The **LSTM-based classifier**, implemented in TensorFlow/Keras, consists of a three-layer stacked LSTM network with 70, 30, and 10 hidden units, respectively. Dropout regularization (rate = 0.2) was applied to the first LSTM layer to mitigate overfitting. The final hidden representation was fed into a fully connected layer with a sigmoid activation function to perform binary classification. The model was trained using the Adam optimizer (learning rate 5×10^{-4}) for 100 epochs with an asymmetric binary cross-entropy loss to penalize misclassification of negative market movements more heavily.

The **Transformer-based classifier**, implemented in PyTorch, projects the same 10-day input sequences into a 64-dimensional embedding space with learnable positional encodings. The representations are processed by a two-layer Transformer encoder (4 attention heads, feed-forward dimension 128). The encoded sequence is mean-pooled and passed to a fully connected classification layer. The model was trained using Adam (learning rate 3×10^{-4}) for 200 epochs with binary cross-entropy loss.

Performance was evaluated on chronological splits using accuracy, precision, recall, F1-score, and Mean Directional Accuracy (MDA). The MDA is defined as:

$$\text{MDA} = \frac{1}{N}\sum_{t=1}^{N}\mathbf{1}\big(\text{sign}(y_t - y_{t-1}) = \text{sign}(\hat{y}_t - y_{t-1})\big),$$

where $\mathbf{1}(\cdot)$ denotes the indicator function. MDA ranges between 0 and 1. Under balanced class probabilities, a value of 0.5 corresponds to random directional guessing, while values above 0.5 indicate predictive skill. In financial time-series forecasting, even moderate improvements above the naive benchmark can be economically meaningful, and sustained MDA levels above 0.6 are generally considered strong out-of-sample performance for liquid equity indices [17, 18].

4. Experiments and Results

4.1. Datasets

- **Stock Price Datasets**

The stock price dataset includes the following variables: Date, Open, High, Low, Close, Adjusted Close, and Volume. Daily returns were computed from closing prices and used for modeling and price movement analysis. Historical data for the NASDAQ Composite and the S&P 500 indices were downloaded from Yahoo Finance[19]. To avoid look-ahead bias, the datasets were split chronologically. For the NASDAQ Composite, the sample was divided into a training set (85%, 2,400 observations, February , 2009–January , 2019) and a test set (15%, 424 observations, January , 2019–June , 2020), with the latter including the onset of the COVID-19 market regime. Similarly, the S&P 500 dataset was split into a training period (85%, June 2009–January 2019) and a test period (15%, January 2019–June 2020), also covering the emergence of the COVID-19 crisis.

- **Financial News Sentiment**

The sentiment dataset contains 13 columns, including the Date column and FinBERT sentiment scores. For NASDAQ, we collected 843,062 news articles from the publicly available NASDAQ News dataset hosted on Hugging Face [20]. The raw news corpus was filtered, processed, and temporally aligned with corresponding stock price observations as part of our data construction pipeline. To focus on relevant market movements, the news was filtered using keywords related to NASDAQ, selecting up to 10 articles per day over the entire period. For the S&P 500, we collected 243,519 news articles from the Hugging Face S&P 500 News corpus [21]. Similar preprocessing, filtering, and alignment procedures were applied to construct the final dataset used for modeling. During preprocessing, we observed incomplete company-level coverage in the raw corpora, which could introduce temporal sparsity in sentiment signals. To ensure consistent daily representation, we collected additional company-related news from the same Hugging Face source when coverage was incomplete. This gap-filling procedure was applied systematically to maintain balanced and continuous sentiment signals across the entire observation period.

The `FinBERT` model was applied to these daily news headlines, and the mean sentiment score for each day was computed. These daily scores, ranging from -1 (strongly negative) to 1 (strongly positive), were used to generate engineered sentiment features for modeling. The mean near zero indicates balanced sentiment over the period.

4.2. Baseline Replication

We first replicated the baseline approach [11] using a single sentiment feature (FinBERT score) with the original LSTM architecture. While regression metrics were strong, directional performance remained limited, as shown in Table 2. In particular, although the baseline model achieved high accuracy (0.955), its MDA was below 0.5 (0.467), indicating poor directional skill and bias toward the majority class.

We then replaced the single daily sentiment score with the selected temporal sentiment features described in Section 3.2. As reported in Table 2, this enhancement improved both regression and directional performance, increasing MDA from 0.467 to 0.524 and accuracy from 0.955 to 0.983. This discrepancy between high accuracy and low MDA further confirms that accuracy alone can be misleading in imbalanced directional forecasting settings.

The global MDA represents the overall proportion of correctly predicted directions, while MDA_+ and MDA_- measure directional accuracy conditional on upward and downward movements, respectively. The

Table 2. Comparison of baseline LSTM with finbert score and LSTM with top temporal sentiment features on NASDAQ.

Model	**MAE**	**MAPE (%)**	**Accuracy**	**MDA**	**MDA_+**	**MDA_-**
LSTM (baseline)	0.00083	21.77	0.955	0.467	0.806	0.257
LSTM + Top Sentiment Features	0.00072	19.55	0.983	0.524	0.725	0.385

strong imbalance between MDA_+ and MDA_- confirms that both model are biased toward predicting upward movements.

4.3. **Classification with Asymmetric Objective**

We evaluate the classification formulation using the proposed asymmetric objective. We retained the same LSTM architecture as the baseline but replaced the output layer with a sigmoid activation. In addition, we introduced a custom asymmetric loss function to penalize incorrect predictions of negative market movements more heavily.

4.3.1. **LSTM Classification Results**

We evaluated the LSTM classification model under two settings to assess the contribution of sentiment feature engineering and the asymmetric objective function.

First, the classification model was trained using only the raw FinBERT sentiment score and standard binary cross-entropy loss, replicating the baseline configuration. As reported in Table 3, the overall directional performance was modest (MDA = 0.515). While this value is slightly above 0.5, it does not necessarily imply strong predictive skill, particularly given potential class imbalance in equity markets.

More importantly, the model exhibited clear asymmetry across classes. The F1-score for positive movements was 0.594, compared to only 0.398 for negative movements. Similarly, recall for the negative class (0.381) was substantially lower than for the positive class (0.613), confirming the model's difficulty in correctly identifying downturns, especially during volatile periods such as the early COVID-19 market shock.

Next, we incorporated the proposed sentiment feature set and applied feature selection, combined with an asymmetric objective function to penalize misclassification of negative movements. As shown in Table 3, this configuration produced consistent improvements across evaluation metrics. The MDA increased from 0.515 to 0.592. Importantly, the F1-score for negative movements improved markedly from 0.398 to 0.526, and negative recall rose from 0.381 to 0.540, indicating substantially better detection of market downturns.

Although the asymmetric loss improved sensitivity to negative movements, higher penalty values reduced performance on positive predictions, reflecting the trade-off inherent in cost-sensitive learning. The parameter λ was selected through hyperparameter experiments by testing values in the range $[0.5, 2.0]$ with increments of 0.25. Model performance was evaluated using class-wise directional accuracy, and $\lambda = 1.25$ was chosen as it provided the best empirical balance between improved detection of negative movements and acceptable performance on positive predictions.

Table 3. Comparison of LSTM classification performance on NASDAQ: baseline vs. enhanced model with sentiment features and asymmetric loss.

Model	**Class**	**Precision**	**Recall**	**F1-score**	**Accuracy / MDA**
Baseline LSTM (FinBERT only)	Positive	0.576	0.613	0.594	0.515
	Negative	0.417	0.381	0.398	
LSTM + Sentiment Features + Asymmetric Loss	Positive	0.640	0.595	0.617	0.592
	Negative	0.513	0.540	0.526	

4.3.2. **Transformer Classifier Performance**

We compared the proposed Transformer classifier with our previous LSTM-based model (see Table 3) to assess improvements in directional forecasting when using sentiment features and an asymmetric loss. The comparison is summarized in Table 4, which presents class-wise precision, recall, and F1-scores for positive and negative market movements, as well as overall accuracy and MDA.

Table 4. Comparison of LSTM and Transformer classification performance on NASDAQ only during the COVID-19 volatile period.

Model	**Class**	**Precision**	**Recall**	**F1-Score**
LSTM + Sentiment Features + Asymmetric Loss	Positive	0.561	0.651	0.603
	Negative	0.559	0.592	0.575
	Accuracy / MDA = ***0.592***			
Transformer + Sentiment Features + Asymmetric Loss	Positive	0.748	0.524	**0.616**
	Negative	0.659	0.839	**0.738**
	Accuracy / MDA = ***0.688***			

Table 4 shows that the Transformer improves overall accuracy and MDA compared to the LSTM. It achieves notably higher recall for negative movements (0.839 versus 0.592 for the LSTM), reflecting improved sensitivity to market downturns. Positive recall is slightly lower (0.524), highlighting the trade-off introduced by the asymmetric loss function. These results indicate that, while the Transformer performs better in detecting negative trends, no model achieves perfect prediction in both directions. Although Table 4 focuses on the NASDAQ dataset, the model was also evaluated on the S&P 500 index to assess generalizability. On the S&P 500, the Transformer achieved an accuracy (and MDA) of 0.661. Precision scores were balanced across classes, with 0.646 for positive movements and 0.675 for negative movements. Recall values were similarly symmetric (0.639 for positive and 0.681 for negative movements), resulting in F1-scores of 0.642 (positive) and 0.678 (negative).

We further compared our Transformer with three recent studies that report directional accuracy on NASDAQ or S&P 500. We applied our model also on S&P 500 during comparable test periods (see Table 5). These studies were selected because they (i) cover the same stock indices, (ii) are closer in time to our evaluation period, and (iii) report overall directional accuracy metrics.

Table 5. Comparison of Transformer with selected recent directional forecasting studies. Models are cited in the related work section.

Study / Model	**Market**	**Directional Accuracy / MDA**
ATFNet + StockMixer [22]	NASDAQ	0.412
Blending Ensemble (PMC8446482) [23]	NASDAQ	0.667
Random Forest + LSTM (Zhong & Hitchcock, 2021) [24]	S&P 500	0.662
LSTM + Sentiment Features + Asymmetric Loss	NASDAQ	0.592
LSTM + Sentiment Features + Asymmetric Loss	S&P 500	0.610
Proposed Transformer	NASDAQ	0.688
Proposed Transformer	S&P 500	0.662

Several observations arise from Tables 4 and 5:

- The Transformer achieves higher overall MDA than LSTM and slightly outperforms the selected recent studies. However, differences in datasets, testing periods, and market conditions make direct comparison challenging.
- Most prior studies report only overall directional accuracy or MDA and do not provide class-wise metrics (positive vs. negative). As a result, their models may achieve high MDA by primarily predicting upward movements, potentially underperforming in detecting negative trends. Our Transformer explicitly maintains strong negative recall, which is particularly relevant for risk management.
- Many high-performing models in the literature are evaluated on relatively stable market periods. By contrast, our evaluation includes the onset of the COVID-19 pandemic, a period of extreme volatility. Maintaining high negative recall during such turbulent times suggests the Transformer is robust to regime shifts.
- While the Transformer performs well, recall for positive movements (0.524) is lower than negative recall, indicating that no model is perfectly balanced. This emphasizes the importance of reporting and analyzing class-wise performance rather than relying solely on aggregated metrics like MDA.

- The comparison highlights that even if other papers report strong directional metrics, their results may not generalize to volatile periods or reflect sensitivity to downward movements. Including positive/negative analysis provides a more realistic and practical assessment for financial forecasting.

These results demonstrate that both overall accuracy and directional balance are crucial for evaluating market prediction models. Our Transformer model shows competitive performance while explicitly addressing the challenge of negative movement detection in volatile market conditions.

5. Conclusion

This paper proposes a directional stock forecasting framework that integrates temporally structured sentiment derived from financial news with historical price series. Unlike traditional regression-based approaches, the prediction task is reformulated as a binary classification problem to better reflect financial decision-making requirements, where anticipating market direction is often more critical than minimizing numerical error.

The methodology combines sentiment representations extracted using FinBERT with temporal feature engineering and feature selection, and evaluates sequential deep learning models, including LSTM and Transformer architectures. In addition, a custom asymmetric objective function is introduced to penalize risky false positive predictions, improving sensitivity to negative market movements while maintaining balanced performance across positive and negative trends.

Experimental results demonstrate that modeling sentiment as a dynamic temporal signal, rather than a single daily observation, contributes to improved directional forecasting. Both LSTM- and Transformer-based models benefit from the proposed formulation, with the Transformer providing competitive performance in capturing complex temporal dependencies and improving directional stability.

Future work will focus on extending the framework to multiple tickers and longer historical periods to improve generalization and robustness across market conditions. Additional feature types, including technical indicators, macroeconomic variables, and richer topic-specific sentiment representations, could further strengthen predictive performance. Exploring adaptive or regime-aware objective functions may also help balance financial risk and opportunity under different market environments. Moreover, generating synthetic financial time series that replicate realistic market behavior could support training when historical data are limited, while zero-shot and transfer learning strategies may enable models to generalize to unseen tickers with minimal retraining. Finally, integrating backtesting procedures and strategy-level evaluation will be essential to assess the practical financial impact of the proposed approach beyond predictive metrics, and hybrid or ensemble architectures may further enhance directional forecasting performance.

Declaration of AI Use

No artificial intelligence (AI) tools were used in the development of the scientific content of this work, including the conception of ideas, methodology design, experiments, or analysis. All research contributions and results presented in this paper are entirely the authors' own. AI-based tools were used solely for language editing purposes, such as improving grammar, clarity, and readability of the manuscript. The authors take full responsibility for the final content of this paper.

References

[1] E. Chong, C. Han, and F. C. Park. "Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies". In: *Expert Systems with Applications* 83 (2017), pp. 187–205.

[2] A. M. Ozbayoglu, M. U. Gudelek, and O. B. Sezer. "Deep learning for financial applications: A survey". In: *Applied Soft Computing* 93 (2020), p. 106384.

[3] P. C. Tetlock. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market". In: *The Journal of Finance* 62.3 (2007), pp. 1139–1168.

[4] J. Kalyani, P. Bharathi, and P. Jyothi. "Stock trend prediction using news sentiment analysis". In: *arXiv preprint arXiv:1607.01958* (2016).

[5] D. Araci. "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models". In: *arXiv preprint arXiv:1908.10063* (2019).

[6] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding". In: *arXiv preprint arXiv:1810.04805* (2018).

[7] T. Fischer and C. Krauss. “Deep learning with long short-term memory networks for financial market predictions”. In: *European Journal of Operational Research* 270.2 (2018), pp. 654–669.
[8] S. Hochreiter and J. Schmidhuber. “Long Short-Term Memory”. In: *Neural Computation* 9.8 (1997), pp. 1735–1780.
[9] A. Thakkar and K. Chaudhari. “Fusion in stock market prediction: A decade survey on the necessity, recent developments, and potential future directions”. In: *Information Fusion* 65 (2021), pp. 95–107.
[10] X. Li, J. Cao, and Z. Pan. “Market impact analysis via deep learned architectures”. In: *Neural Computing and Applications* 31.10 (2019), pp. 5989–6000.
[11] W. Gu, Y. Zhong, S. Li, C. Wei, L. Dong, Z. Wang, and C. Yan. “Predicting Stock Prices with FinBERT-LSTM: Integrating News Sentiment Analysis”. In: *arXiv preprint arXiv:2407.16150* (2024).
[12] W. Zhang and F. Liu. “Stock Trend Forecasting with Deep Learning and Sentiment Analysis on NASDAQ”. In: *Journal of Financial Data Science* (2023). Directional Accuracy: 50%, Recall Positive: 75%, Recall Negative: 25%.
[13] X. Li and Y. Chen. “Deep Neural Networks for SPY Stock Movement Prediction with Sentiment Features”. In: *Applied Intelligence* (2022). Accuracy: 88%, MDA: 52%, Recall Positive: 70%, Recall Negative: 34%.
[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. “Attention is All You Need”. In: *Advances in Neural Information Processing Systems*. 2017.
[15] L. Breiman. “Random Forests”. In: *Machine Learning* 45.1 (2001), pp. 5–32.
[16] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. Cambridge, MA: MIT Press, 2016. ISBN: 9780262035613.
[17] R. S. Tsay. *Analysis of Financial Time Series*. 3rd. Hoboken, NJ, USA: John Wiley Sons, 2010.
[18] J. Y. Campbell and S. B. Thompson. “Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?” In: *Review of Financial Studies* 21.4 (2008), pp. 1509–1531. DOI: 10.1093/rfs/hhn050.
[19] Yahoo Finance. *Historical Market Data*. https://finance.yahoo.com/.
[20] Hugging Face. *NASDAQ News Dataset (FNSPID_ nasdaq)*. https://huggingface.co/datasets/sabareesh88/FNSPID_nasdaq.
[21] Hugging Face. *S&P 500 News Dataset (FNSPID Stock News)*. https://huggingface.co/datasets/Zihan1004/FNSPID/tree/main/Stock_news.
[22] Anonymous. “Research on Deep Learning Model for Stock Prediction (ATFNet + StockMixer)”. In: *Scientific Reports* (2025). URL: https://www.nature.com/articles/s41598-025-14872-6.
[23] Anonymous. *Blending Ensemble for Stock Direction Prediction*. PubMed Central. 2021. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8446482/.
[24] Y. Zhong and J. Hitchcock. “Random Forest + LSTM for Stock Directional Forecasting”. In: *Statistics, Optimization Information Computing* (2021). URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMCXXXXXX/.