Stock Price Prediction based on Context-Aware Language Models

Jun Li **Beining Jin** **Linqiao Yang** **Yichuan Zhang**
University of Michigan, Ann Arbor
{lju, jinbein, joelyang, yichuanz}@umich.edu

Abstract

This paper proposes a stock price prediction model that integrates qualitative financial news data with quantitative historical price signals, reframing prediction as a sentiment-based classification task. It utilizes an Average True Range (ATR)-based approach for labeling market sentiment and FinBERT, a finance-specific language model, to enhance alignment with market trends. Preliminary experiments indicate parameter values K_1=1.2 and K_2=0.4 as optimal. However, initial evaluations reveal limited accuracy from the untuned FinBERT model, emphasizing the necessity for further fine-tuning to improve predictive performance.

1 Introduction

1.1 Problem description and motivation

The global stock market represents an immense financial ecosystem, with a total capitalization exceeding $100 trillion, reflecting its critical importance to global economic stability and prosperity. Despite its enormous scale, stock prediction remains an inherently complex and challenging task due to the stochastic nature of financial markets and the multitude of factors influencing stock movements. Traditional stock prediction models often rely on Long Short-Term Memory (LSTM) networks and Support Vector Machines (SVMs). While LSTMs are well-suited for time-series analysis and SVMs effectively handle high-dimensional data, they still have certain limitations in stock price prediction.

Today, a single social media post from a public figure can trigger billion-dollar fluctuations in global markets within minutes. As markets evolve, they are increasingly shaped by the collective behavior of investors. Investor sentiment, often reflected in the news, serves as a proxy for market expectations.[1] Given the substantial impact of market sentiment on volatility and the difficulty to quantify the impact of news on stock prices, this study aims to treat stock market prediction as a classification task.

By integrating Average True Range (ATR), an indicator used to quantify the average amplitude of price fluctuations[2] with sentiment features derived from the BERT for Financial Text Mining (FinBERT), a market -sentiment based classification approach is developed to predict short-term market trends and develop trading strategies. The core challenge is effectively combining ATR indicator with FinBERT sentiment features to enhance prediction accuracy, algorithmic trading strategies.

1.2 Impact

Successfully addressing the integration challenge between qualitative and quantitative data can lead to profound impacts on financial markets. Enhanced predictions from this combination can help create more profitable and reliable trading strategies, giving traders and investors useful insights

EECS 545: Machine Learning, Winter 2025. University of Michigan. (Latex Template borrowed from NIPS 2021)

for making decisions. Also, using news sentiment along with price data can greatly improve risk management by providing early warnings about possible market changes or financial crises.

In addition, improved predictive models can help market regulators quickly find and handle unusual trading behaviors based on market news. Overall, sentiment classification analysis can lead to more efficient and stable markets, boosting investor confidence and supporting the health and sustainability of the global financial system.

2 Related Work

In the past, stock price predictions are always based on time-series model like ARIMA model [3] or neural network model like RNN, LSTM and CNN [4]. These methods mainly focus on the stock price data and can not capture the substantial influence of market news data. Sentiment analysis in financial markets emerged from the growing recognition that market psychology plays a crucial role in market movements[5]. Recent advances in Natural Language Processing (NLP) have enabled models to incorporate text-based financial insights, significantly improving prediction accuracy. Studies have shown that social media sentiment (e.g., Twitter, Reddit) influences short-term market movements. As a result, several models have combined NLP techniques with traditional prediction models to enhance accuracy.

For text representation, models such as BERT[6] have been used to outperform traditional sentiment analysis methods in extracting market-relevant signals. BERT generates representations of words or sentences in relatively low-dimensional spaces, allowing for a more contextual understanding of financial news. A fine-tuned version of BERT, FinBERT, specifically trained on financial text, has demonstrated even greater accuracy in capturing market-relevant insights[7].

Other studies have integrated NLP with time-series models. Self-Organizing Fuzzy Neural Networks (SOFNN)[8] have been used to analyze sentiment from financial news. When incorporated into a hybrid SOFNN-based model, sentiment analysis significantly improved stock price prediction accuracy compared to traditional methods. Furthermore, other models[9] have leveraged Recurrent Neural Networks (RNNs) and Gated Recurrent Units (GRUs) alongside NLP-derived sentiment scores, demonstrating that sentiment information effectively complements traditional financial indicators in stock forecasting.

3 Proposed methods

To enhance financial sentiment analysis by incorporating market context, we propose a method that integrates financial news with market movement indicators. Our approach refines the sentiment classification process by leveraging the Average True Range (ATR) metric to construct sentiment labels based on stock price fluctuations. We then fine-tune FinBert using these enriched labels to acquire embeddings and adopt transformer to make predictions with these embeddings. The figure below shows the architecture of our project.

3.1 Dataset Collection & Preprocessing

For market stock price data, we use python library `yfinance`. The `yfinance` provides daily open, close, high, low price and trading volume, also known as 'OCHLV' data.

For Financial news data, We use the open FNSPID Financial news Dataset on Hugging Face[10]. This dataset contains 6537 Intraday news, which are real-time news on market price direction of current day, and about 90,000 industry development news for the related sector. Each news included metadata such as title, date released, full article text, and corresponding summaries produced by LexRank algorithm. We concatenate title and LexRank summery for further embedding stage, since original article includes more token than FinBERT's context window can handle.

Then we pair up the price and news from the same day.

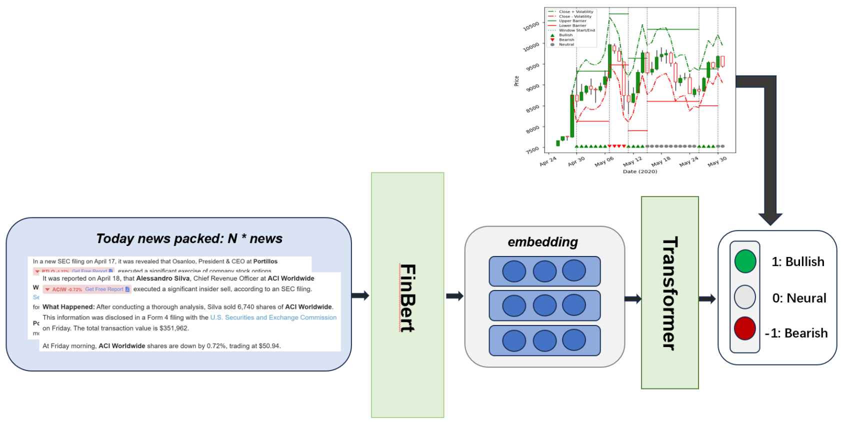

Figure 1: Utilized FinBert to embed one day's news, then apply transformer to aggregate.

3.2 Market-Aware Labeling Process

The core innovation of our approach lies in the market-aware sentiment labeling process, which links financial news sentiment with actual stock movements. This process consists of three main steps:

3.2.1 ATR-Based Market Signal Construction

ATR (Average True Range) is an indicator used to measure how much the price of an asset moves (its volatility). When the price moves a lot, ATR goes up, causing the upper and lower ATR bands around the price to widen. Traders use these bands to gauge market sentiment: if the price rises above the upper band, sentiment is positive (+1); if it drops below the lower band, sentiment is negative (-1).

To calculate ATR, traders first find the True Range (TR). TR is the largest of three possible distances: the distance between today's highest and lowest prices, today's highest price to yesterday's closing price, and today's lowest price to yesterday's closing price. ATR is then simply the average of these TR values over a set number of days (usually 14, but 5 days were chosen here).

$$TR_t = \max\{|High_t - Low_t|, |Close_{t-1} - High_t|, |Close_{t-1} - Low_t|\}$$

$$ATR_t(N) = \frac{1}{N}\sum_{i=1}^{N} TR_{t-N+1}$$

3.2.2 Best hyper-parameter selecting

Next, the strategy was tested using two key parameters, K_1 and K_2. K_1 and K_2 decides the width of our band. Choosing a proper K_1 and K_2 is extremely important. If they are too small, the labels are largely determined by noise in the market, and it makes hard for model to understand the news during traning processs. Proper K_1 and K_2 also contributes to balance of lables in our data.

We adjust these 2 parameter by observing how they affect trading performance on daily S&P 500 data from 2020 to 2024. These parameters set how far the ATR bands are from the price. Testing showed that smaller K_2 values increased volatility, while larger K_1 values reduced volatility and risk. The best overall results (higher returns and lower risks) were found when K_1 was relatively large and K_2 was smaller.

After reviewing various combinations, the recommended parameters are K_1 = 1.2 and K_2 = 0.4, as they balance return and risk effectively. For future improvements, it's suggested to make sentiment labels smarter by adding extra conditions when deciding to open a trade and continuously adjusting sentiment during the trade. This would better capture how strong and lasting market trends actually are.

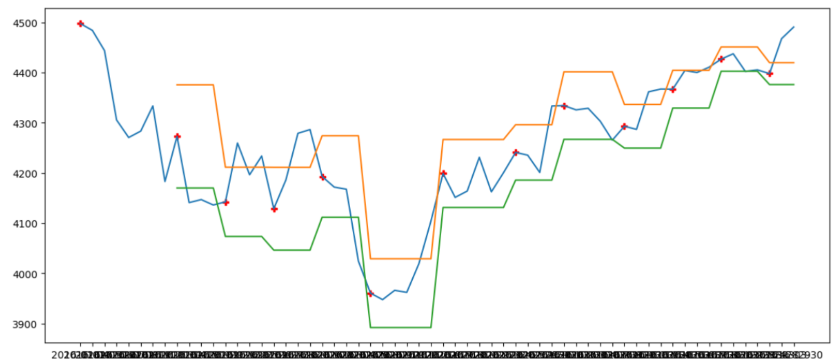

Figure 2: Example of upper ATR bands and lower ATR bands on s&P 500 stock index. The blue line shows path of market stock price; orange line shows the upper band; green line shows the lower band. We use the past market price to compute the uppper and lower band for 1 week future time window. If the market hits the upper band first, all the news in this time window will be labeled as 1; else if it hits lower band first, it will be labeled as -1; else 0.

3.3 FinBERT embedding & Transformer

We adopt a step-wise architecture that leverages FinBERT embeddings and Transformer with ATR-based technical indicators to construct sentiment-aware trading signals.

We begin by utilizing the pre-trained FinBERT model to convert financial news text into contextualized embeddings. These embeddings are then passed through a Transformer-based architecture with attention mechanisms to capture nuanced sentiment signals relevant to stock movements.

To incorporate market volatility, we correlate the sentiment outputs with the Average True Range (ATR) indicator. Specifically, we define sentiment labels (Positive, Neutral, Negative) based on whether the subsequent price movement crosses ATR-derived volatility thresholds. This integration allows us to align news-driven sentiment with technical price boundaries and enhance predictive accuracy.

We train the FinBERT-Transformer model on labeled historical data, using the extracted news embeddings and corresponding ATR-based labels. The model is optimized based on sentiment classification accuracy. This enables the model to generate actionable signals that align with a rule-based trading system.

By combining pre-trained language understanding with volatility-aware labeling and transformer-based sequence modeling, our approach enables precise mapping from financial news to directional trading decisions.

4 Result

4.1 Preliminary result of ATR indicator

Considering that the data frequency is '1-day', Using the market data from January 1, 2020 to January 1, 2024, we calculated the indicators that investors are more concerned about, observed the impact of different combinations of K_1 and K_2 on the various evaluation indicators of the strategy, and found the optimal combination in the sample as the strategy parameter used outside the sample. In actual trading, the length of the training set and the test set depends on specific needs. If the optimal solution parameters obtained in the above process do not perform well in actual returns, it means that the current market is not compatible with the historical market, and it is necessary to adjust the parameters in time or reduce the strategy opening rights. We take N=5, and the performance of the strategy on the NASDAQ index when K_1 and K_2 are constantly changing is as follows:

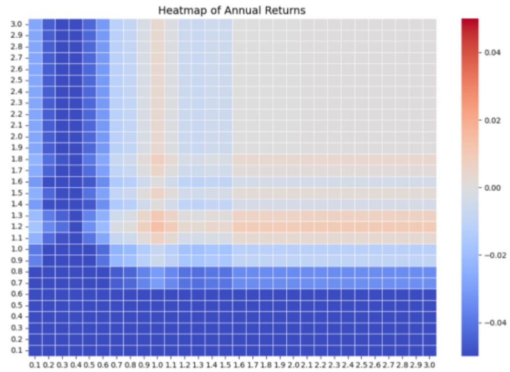

Figure 3: Heatmap of Annual Return

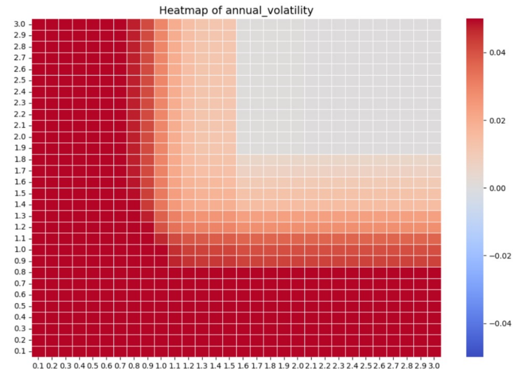

Figure 4: Heatmap of Annual Volatility

The analysis shows that parameters K_1 and K_2 significantly influence the annualized return, volatility, Sharpe ratio, maximum drawdown, and winning rate. Specifically,

$$K_2$$

exhibits higher sensitivity, strongly affecting returns, with larger K_1 values providing greater stability. Increased K_1 reduces volatility, while smaller K_2 values lead to higher market fluctuations. A combination of higher K_1 and lower K_2 optimizes the Sharpe ratio and enhances the winning rate, though their impact on maximum drawdown lacks clear patterns—generally lower drawdowns occur when K_1 exceeds 2.0. Considering return, risk, and the market characteristic of being "easy to fall but difficult to rise," the optimal parameter combination within the sample is recommended as K_1 = 1.2 and K_2 = 0.4, balancing returns with manageable risk and limiting potential drawdown.

4.2 Baseline experiment

As outlined previously, we leverage FinBERT to extract sentiment tokens from financial news and use these embeddings to predict next-day stock-price fluctuations. First, we establish baselines using traditional models—Random Forest, XGBoost, and an LSTM—each trained on the daily average FinBERT-derived token embeddings. Next, we explore a single-headed Transformer architecture, addressing varying daily token counts through three strategies: (1) averaging all token embeddings of one day's news data, (2) randomly sampling up to 500 daily tokens of news data embedding to mitigate overfitting, (3) combining daily embeddings with past three-day prediction signals. Table below presents the evaluation results for all configurations.

Table 1: Model Results

Methods	**Validation Accuracy**
Random Forest	0.5315
XGBoost	0.5545
LSTM	0.6147
Transformer (average daily input)	0.5813
Transformer (randomly choosing)	0.5252
Transformer (with past signal)	**0.8460**

The Transformer model that combines FinBERT-derived sentiment embeddings with the three most recent true fluctuation signals achieves by far the best performance, yielding 0.846 validation accuracy. In contrast, our LSTM baseline attains 0.612, while traditional machine-learning methods trained solely on the same token embeddings—Random Forest and XGBoost—reach just 0.532 and 0.555, respectively. Variants of the single-headed Transformer that rely exclusively on embedding information perform comparably poorly: averaging all daily embeddings produces 0.581 accuracy, and concatenating each day's embeddings without any price signals drops further to 0.525. These results underscore that fusing textual sentiment with recent price dynamics is critical for accurate next-day ATR fluctuation prediction.

5 Conclusion

In our work, we propose a two-stage pipeline to predict daily stock movement based on textual information. In particularly, we extract contextualized embeddings using FinBERT and train a Transformer-based classifier to aggregate infomation for downstream prediction. To label each trading day, we adopt a technical analysis-inspired approach to incorporate market volatility.

Our method achieves a 22% improvement in return over a baseline that uses only past price data, demonstrating the effectiveness of deep sequential modeling for financial text. Our work demonstrate the potential of language models to extract dense context representation and finance prediction.

Futrue research can focus on 2 directions. First, by leveraging return-aware and adaptive signal learning moving beyond binary classification, future models can incorporate return-based objectives—such as excess return or risk-adjusted return—into training. This allows the model to directly learn representations aligned with financial reward. Moreover, replacing fixed ATR thresholds with learnable volatility parameters, optimized end-to-end, would enable the system to adaptively calibrate its signal generation based on evolving market conditions. Second, reinforcement learning agents can be introduced to optimize signal thresholds, timing, and position sizing in a sequential setting. Additionally, incorporating multi-modal inputs, such as technical indicators and market regime features, may enhance model robustness and allow context-aware forecasting across varying market environments.

6 Author Contributions

All authors contributed equally to this work. Beining Jin conducted the ATR indictor classification, Linqiao Yang conducted the FinBert embedding , Yichuan Zhang conducted the baseline model and Jun Li conducted the Transformer model. The paper was written by all authors together. All authors discussed the results and implications and commented on the manuscript at all stages.

Task	Beining Jin	Jun Li	Linqiao Yang	Yichuan Zhang
Model Design	✓	✓	✓	✓
Code Writing	✓	✓	✓	✓
Paper Writing	✓	✓	✓	✓

References

[1] How elon musk's twitter activity moves cryptocurrency markets. *Technological Forecasting and Social Change*, 186:122112, 2023. 1

[2] Cohen Gil. Trading cryptocurrencies using algorithmic average true range systems. *Journal of Forecasting*, 42, 08 2022. 1

[3] Adebiyi A. Ariyo, Adewumi O. Adewumi, and Charles K. Ayo. Stock price prediction using the arima model. In *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, pages 106–112, 2014. 2

[4] Sreelekshmy Selvin, R Vinayakumar, E. A Gopalakrishnan, Vijay Krishna Menon, and K. P. Soman. Stock price prediction using lstm, rnn and cnn-sliding window model. In *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1643–1647, 2017. 2

[5] Andrew W Lo. The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *Journal of Portfolio Management, Forthcoming*, 2004. 2

[6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019. 2

[7] T Swathi, N Kasiviswanath, and A Ananda Rao. An optimal deep learning-based lstm for stock price prediction using twitter sentiment analysis. *Applied Intelligence*, 52(12):13675–13688, 2022. 2

[8] Jaydip Sen and Sidra Mehtab. A robust predictive model for stock price prediction using deep learning and natural language processing. July 2021. 2

[9] Ryo Akita, Akira Yoshihara, Takashi Matsubara, and Kuniaki Uehara. Deep learning for stock prediction using numerical and textual information. In *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, pages 1–6. IEEE, 2016. 2

[10] Zihan Dong, Xinyu Fan, and Zhiyuan Peng. Fnspid: A comprehensive financial news dataset in time series, 2024. 2