UNEC Journal of Computer Science and Digital Technologies
Volume 1, № 1, 2025, pp. 27-38
https://doi.org/10.30546/UNECCSDT.2025.01.027

Section:
Artificial Intelligence &
Machine Learning

# Evaluating hybrid deep learning models for financial market trading

**Guljan Mammadova[1*], Eltun Babirzada[2]**

*[1*]Department of Business Administration, Baku Higher Oil School, Baku AZ1003, Azerbaijan*
*[2]Department of Petroleum Engineering, Baku Higher Oil School, Baku AZ1089, Azerbaijan*
*0009-0005-5417-6575, guljan.mammadova.std@bhos.edu.az*
*0009-0008-1810-439X, eltun.babirzade.std@bhos.edu.az*

**Abstract**

The prices of stocks are hard to predict due to the complex interplay of qualitative sentiment and quantitative indicators. This article analyzes 2 strategies for feature extraction which affect the quality of hybrid deep learning models. To conduct the experiment, the New York Times articles were web scraped, and the extracted sentiment data were combined with S&P500 historical data. Two hybrid models were developed: one using VADER sentiment analysis followed by a CNN to extract refined sentiment features before passing them to an LSTM, and another using FinBERT to generate contextual sentiment embeddings directly fed into an LSTM. Interestingly, the VADER-CNN-LSTM model achieved slightly better predictive performance, suggesting that lexicon-based methods, when combined with effective feature extraction layers like CNNs, can be competitive in small-data scenarios. Although FinBERT is designed to capture nuanced, domain-specific sentiment signals, its advantages may be underutilized without sufficient data volume. This highlights a key insight: the success of sentiment-enhanced forecasting models depends not only on model architecture, but also on the quantity and structure of the available data.
`

***Keywords***: stock price prediction, deep learning, financial news, hybrid models

| *Received:* | *Revised:* | *Accepted:* | *Published:* |
|---|---|---|---|
| *29/05/2025* | *03/06/2025* | *06/06/2025* | *14/06/2025* |

## 1. Introduction

Forecasting stock prices is a significant topic in financial markets trading and quantitative finance, having very profound implications for investment strategies, risk management, and market efficiency. Despite its theoretical limitations rooted in the Efficient Market Hypothesis, real-world patterns have encouraged researchers to explore whether relevant external data sources, particularly those reflecting investor sentiment, can help enhance prediction accuracy. The idea is inspired by the persistent gap between the theoretical framework of the Efficient Market Hypothesis and empirical observations of market anomalies caused by behavioral factors such as investor sentiment [1, 2]. The growing availability of high-frequency and high-dimensional alternative datasets such as social media posts, financial news articles, and search engine trends has catalyzed usage of machine learning and deep learning models for designing hybrid forecasting pipelines, as seen in papers linking sentiment from news, social media, and search trends to market movements [3 -7].

In this research, we propose and evaluate two sentiment-driven hybrid models that combine natural language processing methods with deep learning architectures: (1) VADER-

CNN-LSTM and (2) FinBERT-LSTM. Both models aim to forecast the closing price of the S&P 500 using a combination of sentiment signals extracted from daily financial news and numerical stock data.

At the heart of our motivation lies the hypothesis that financial news sentiment influences short-term investor behavior and, subsequently, market prices. Specifically, we leverage daily textual data from the New York Times, capturing the first sentence of the top 10 financial, business, or technology-related articles for each trading day between October, 2020, and September 2022. These sentences, often referred to as lead sentences in media terminology, function as succinct summaries of news content and serve as effective proxies for capturing the sentiment communicated in financial reporting. Unlike headlines, which are carefully crafted for attention-grabbing purposes, lead sentences are more informative and narrative-driven, offering a clearer signal of the story's tone. This choice eliminates concerns about the noise and potential sensationalism present in headlines and aims for a reliable sentiment proxy [8].

In our first hybrid model (VADER-CNN-LSTM), we begin by extracting sentiment labels using the VADER model. VADER is a rule-based sentiment analysis tool and it is mainly used for texts derived from social media but has proven effective in short, general-purpose English texts as well. After computing compound sentiment scores from the VADER output, we aggregate the sentiment of each day's news to form a single numerical score and corresponding label (positive, negative, or neutral). These sentiment labels are used as ground truth to train a CNN model for classification, which, once trained, generates probabilistic sentiment outputs that better capture the nuances in financial text.

Our second hybrid model (FinBERT-LSTM) instead uses a pre-trained FinBERT model to generate sentiment scores directly. FinBERT is a domain-focused BERT which has been fine-tuned on financial texts. It is particularly suitable for parsing and interpreting the semantic subtleties of business-related language, which general-purpose models often misclassify. It addresses the core challenge of domain adaptation in financial text analysis, while general-purpose tools tend to underperform [9]. The daily average FinBERT sentiment scores (mean of the top 10 articles) are then paired with normalized stock price sequences and used to predict future price movements using an LSTM model.

Both approaches culminate in the use of LSTM networks that are mainly utilized in time series problems, and it has the ability to maintain long-term dependencies and patterns of sequential data. By combining these architectures, we attempt to quantify the marginal predictive power of sentiment information when incorporated into numerical forecasting pipelines.

Below, we will delve into the technical construction of the VADER-CNN-LSTM model, providing detailed descriptions of preprocessing, architectural choices, and evaluation metrics, followed by an analogous breakdown for the FinBERT-LSTM pipeline. Through comparative experimentation, we assess the effectiveness of each model in the prediction of the S&P 500 index and explore how sentiment features extracted from financial news articles contribute to predictive performance. The central research question driving this work is whether incorporating these carefully extracted sentiment features provides statistically significant marginal predictive power beyond traditional numerical data alone, a finding with tangible implications for quantitative trading strategies.

## 2. Literature review

Forecasting stock prices has long attracted interest from both academic researchers and financial practitioners. Numerous studies have explored the integration of traditional market data with textual sentiment signals to improve predictive accuracy.

Researchers have investigated the prediction accuracy of different deep learning models in the Tehran Stock Exchange, utilizing five models including two hybrid models such as CNN-LSTM, CNN-GRU and three single models CNN, LSTM, GRU. The hybrid CNN-LSTM was compared against LSTM, GRU, CNN, combined CNN-GRU models over 5 firms. The combined CNN-LSTM model demonstrated the best performance overall, while the CNN model showed comparable results to the combined models in some cases [10].

A group of researchers focused on market forecasting with the help of deep learning models by specifically analyzing the KOSPI market in South Korea. Models used were linear ones, for example, Fama-French, and nonlinear models such as artificial neural networks; specific comparisons include RBM (Restricted Boltzmann Machine), AE (Autoencoder), and PCA (Principal Component Analysis) [11]. Related studies indicated that some linear models outperformed nonlinear models [12, 13]. Conversely, other studies found nonlinear models outperforming linear models [14, 15].

A total of nine deep learning neural network models have been tested for making stock market predictions using a standardized dataset (S&P500) from 2000 to 2017, ensuring a balanced dataset for fair comparison. These were LSTM, GRU, CNN, DQN, RNN, ESN, DNN, RBM, DBN, all using the identical hyperparameters to get unbiased outcomes. The analysis indicates that model enhancements and hyperparameter tuning significantly impact prediction performance, although specific models outperform others in different scenarios, highlighting the need for further research to develop robust stock prediction models [16].

In an attempt to predict prices using news articles and stock data taken from the Nikkei newspaper (2001-2008) for 10 companies, scholars have proposed a model (using Paragraph Vector and LSTM). Their baseline methods included Bag-of-Words (BoW), numerical-data-only methods, and other models like SVR, MLP, Simple-RNN. The model outperformed the baseline methods in terms of total earnings and capturing time series influences, demonstrating superior predictive capabilities when considering multiple companies within the same industry [17].

Furthermore, the relationship between public sentiment derived from Twitter data and stock market predictions, specifically focusing on the DJIA index has been investigated [18]. For the study, linear regression, logistic regression, support vector machine (SVM), self-organizing fuzzy neural networks (SOFNN) have been proposed. Among all, linear regression achieved the best mean absolute percentage error (MAPE) of 7.05% and a direction accuracy of 71.11%. SOFNN had a MAPE of 9.22% with a direction accuracy of 64.44%. Logistic regression and SVM showed lower performance with consistent direction accuracy of 60%.

Another study examines the predictability of stock values via various statistical methods, specifically focusing on the Tehran Stock Exchange over a 10-year period (2009-2019). The research compares tree-based models (Decision Trees, Bagging, Random Forest, Adaboost, Gradient Boosting, XGBoost) and deep learning methods (RNN, LSTM) for regression problems. Among the models, Adaboost and Gradient Boosting showed strong performance, with Adaboost achieving the lowest error metrics, while LSTM outperformed all other models in terms of prediction accuracy [19].

Additionally, stepwise regression analysis, differential evolution-based fuzzy clustering, and a fuzzy inference neural network are built by scholars for prediction of stock prices [15]. Their hybrid model was compared against multi-layer feed-forward neural network, generalized regression neural network, probabilistic neural network, linear regression. This model outperformed all compared ones, achieving the lowest RMSE of 0.9834, which means it was more accurate in its predictions.

Related work evaluates the performance of FinBERT on financial datasets, including sentiment analysis, question answering, and sentence boundary detection. The scholars have compared FinBERT (both BASE and LARGE versions) vs. Vanilla BERT, BERT-task, and state-of-the-art models (e.g., CUKG-TongJi, eLabour). As a result, FinBERT consistently outperformed all baseline models across tasks, achieving state-of-the-art results in financial sentiment analysis and question answering, with FinBERTLARGE showing significant improvements over BERT and other models in metrics like nDCG and MRR [20].

As a hybrid framework, HAN-SPL has been compared against various baseline models over a one-year trading simulation (May 2016 to March 2017). Baseline models were One-RNN, News-RNN, Temp-ATT (Temporal Attention RNN), News-ATT (News Attention RNN), Multi-layer Perceptron (MLP). HAN-SPL outperformed all baseline models, achieving the highest accuracy (0.478) and annualized return (0.611) when investing in the top 40 stocks, significantly surpassing the market performance (0.04) [21].

At-LSTM model has also been tested on financial news data from Reuters and Bloomberg, focusing on predicting the S&P 500 index and individual stock prices from October 2006 to March 2018. The compared models were SVM, Bag-At-LSTM, WEB-At-LSTM, Ab-At-LSTM, Doc-At-LSTM, Tech-At-LSTM, CNN-LSTM, E-NN, EB-CNN, KGEB-CNN, and At-LSTM. Finally, KGEB-CNN model outperformed others with a maximum accuracy of 66.93%, while the At-LSTM achieved a maximum accuracy of 65.53%, indicating that deep learning models generally performed better than traditional SVM [22].

Also, researchers worked on a Deep Q-Network (DQN) model trained on stock price patterns from the US market and tested across 31 countries over a 12-year period. Having been compared against traditional market-neutral portfolios and random portfolios, DQN consistently outperformed both the market-neutral and random portfolios, achieved higher returns and demonstrated the ability to identify profitable trading actions [23].

By using FiQA sentiment scoring and Financial PhraseBank datasets, FinBERT has also been tested on financial sentiment analysis tasks [24]. In this paper, FinBERT has been compared with ULM-Fit and ELMo, both pre-trained models aiming for the analysis of financial sentiment. FinBERT achieved prominent outcomes on both of the datasets, outperforming ULM-Fit and ELMo in sentiment classification tasks.

By comparing TDQN model against traditional trading tactics, including buy/hold, simple moving average, trend following, and mean reversion, research found out that the TDQN model consistently outperformed the traditional strategies, achieving higher Sharpe ratios across multiple stocks, indicating superior risk-adjusted returns [25].

Another review focusing on various deep reinforcement learning approaches compared Deep Q-Network which yielded better results than baseline models such as buy and hold, Sign, and MACD signal, actor-only algorithm (PG) and A2C algorithm which was tested alongside DQN for performance comparison, GDPG (actor-critic strategy) that combined Q-Network with a policy network, achieving more stable risk-adjusted returns than the Turtle trading strategy [26]. As a result, DQN and GDPG demonstrated superior performance, with DQN achieving annual returns of around 22-23%, while other reviewed agents exceeded 60% annual return.

Tests of various stock prediction models across four datasets: CSI 300, CSI 500, S&P 500, and NASDAQ 100 have been done. They built traditional and deep learning models: BLSW, CGM, GRU, LSTM, ALSTM, transformer, and more complex models: TRA, CTTS, A2C, DDPG, PPO, TD3, SAC, AlphaStock, DeepPocket, DeepTrader, FactorVAE, THGNN. The proposed LSR-IGRU model outperformed all baseline models in terms of annualized return rate, annualized sharpe ratio, calmar ratio, and information ratio, demonstrating the effectiveness of leveraging long and short-term relationships in stock prediction [27].

Research compares the traditional Fama–French five-factor model (which includes market, size, value, profitability, and investment factors) with a sentiment-augmented model that incorporates sentiment metrics derived from FinBERT [28]. The sentiment-augmented model outperforms the traditional Fama–French model in explaining abnormal returns, particularly during periods of market volatility, highlighting the significance of behavioral factors in asset pricing.

FinBERT (domain-specific model), GPT-4 (general-purpose model), and Logistic Regression (traditional model) have been tested on financial data. According to the results, logistic regression performed more satisfactorily than both FinBERT and GPT-4, achieving an accuracy of 81.83% and ROC AUC of 89.76%. FinBERT achieved an accuracy of 63.33%, while GPT-4 had the lowest accuracy at 54.19%. This information can be used to highlight the effectiveness of traditional models in financial sentiment analysis compared to advanced NLP models [29].

Assessment of the predictive power of a combined FinBERT-LSTM model against traditional LSTM approaches specifically analyze their effectiveness on six US-listed stocks: Microsoft, Visa, Citibank, Meta, Autodesk, and JP Morgan Chase. The Fin-BERT-LSTM model outperformed traditional LSTM models for four out of the six stocks, achieving a notable 38% reduction in Mean Squared Error for Citibank and Meta [30].

It is noteworthy to highlight the study that focuses on comparing various deep learning and statistical models for stock price forecasting using S&P 500 index data from Dec 2015 to Jan 2018. Statistical models were ARIMA, moving averages, whereas deep learning models were recurrent neural network, long short-term memory (LSTM), convolutional neural network (CNN), and Full CNN. LSTM model achieved the lowest MAE of 1.18, outperforming all other models. Full CNN outperformed pre-process CNN in terms of accuracy. While effective, ARIMA did not match the performance of deep learning models [31].

Evaluation of sentiment analysis models on a collected dataset using 10-fold cross-validation compares the CNN model against four traditional models: logistic regression, support vector machine (SVM), recurrent neural network, and long short-term memory (LSTM). The CNN model outperformed all other models, achieving the highest average precision and recall [32].

One of the articles focused on building CNN-LSTM: The CNN model designed for sentiment analysis was combined with the LSTM model for prediction. The proposed hybrid model consistently outperformed traditional models, showing higher accuracy and reliability in price forecasting [33].

In one of the papers investigated, the performance of three models has been assessed over a five-year period: CNN, GRU, and a fused CNN-GRU framework. The CNN-GRU model significantly outperformed both CNN and GRU individually, achieving precision of 84.32% and F1 score of 0.87, indicating enhanced efficacy in sentiment analysis and risk prediction [34].

Finally, traditional CNNs, recurrent CNNs, and attention-based CNNs have also been evaluated against each other [35]. The attention-based CNN model outperformed traditional and recurrent CNNs in terms of accuracy and efficiency in classifying sentiment.

While numerous studies have employed sentiment analysis to improve stock market forecasting, there remains limited consensus on which sentiment extraction methods perform best, particularly in small or medium-sized datasets.

## 3. Data

*Source 1: Yahoo Finance - S&P 500 Index*

To support time-series forecasting, we collected historical price data for the S&P 500 index from Yahoo Finance. The dataset spans from early October, 2020, to the end of September, 2022. The primary feature used for forecasting was the daily closing price, a widely accepted indicator of stock performance. Additional variables such as open, high, low, etc. were also recorded, though the forecasting model ultimately concentrated on closing prices as the target variable.

*Source 2: New York Times (NYT) News Headlines*

To extract qualitative signals reflective of investor sentiment, we retrieved business-relevant news headlines from the New York Times website. The scraping process focused on the "Business," "Finance," and "Technology" sections. For each trading day, the first sentence of the top 10 articles was extracted, as this typically summarizes the article's core content. In cases of insufficient article volume, fewer headlines were processed, maintaining integrity rather than injecting irrelevant data.

## 4. Data Preprocessing

*Numerical Data:*

Numerical stock price data underwent preprocessing. All numerical features were normalized using MinMaxScaler to constrain values within the [0, 1] range. This step ensured that model convergence was not adversely impacted by differing feature scales, particularly when combining numerical and sentiment-derived features.

*Text Data:*

For the VADER-based model, text preprocessing involved tokenizing headlines, converting text to lowercase, removing punctuation and stopwords, and formatting text for VADER input. Since VADER is a lexicon-based model, these steps preserved critical sentiment-bearing words while eliminating noise.

For FinBERT, a transformer-based model, preprocessing followed its tokenizer specifications. Headlines were first lowercased (as required), then tokenized using FinBERT's pretrained tokenizer. This step preserved contextual integrity, allowing FinBERT to analyze sentences holistically

*Alignment of Sentiment and Price Data:*

One of the key challenges was temporal alignment between textual and numerical data. Since financial markets react to news during or shortly after publication, we mapped each day's average sentiment (from the 10 extracted headlines) to that day's closing price. This created a sentiment-aware sequence of inputs where each sentiment score was assigned to the corresponding trading day. Only days with both stock data and sentiment data were retained to maintain synchronization

## 5. Models' Architecture

Hybrid Model 1: CNN-LSTM with VADER Sentiment Analysis

*Part 1: Sentiment Processing via VADER and CNN*

The first component of the hybrid model involves processing qualitative sentiment using VADER and a Convolutional Neural Network (CNN). Each day's 10 preprocessed headlines were passed through VADER, which assigned a compound sentiment score to each sentence, scaled between -1 (very negative) and +1 (very positive). These scores were then used to generate a 10-dimensional sentiment vector for each trading day.

This vector was fed into a CNN architecture designed to extract higher-level sentiment features. The CNN consisted of:

- Embedding Layer: Transformed sentiment tokens into dense 128-dimensional vectors. Setting trainable=True allows the embeddings to be updated during training.
- Convolutional Layer: This layer applies 64 one-dimensional convolutional filters with a kernel size of 3. This enables the model to detect trigrams or local patterns in the sentiment sequences. The ReLU activation introduces non-linearity.
- MaxPooling Layer: This downsamples the output of the convolution layer by a factor of 2, reducing the spatial dimensions and emphasizing the most prominent features.
- Dropout Layer: Applied twice in the architecture, this layer disables 50% of the neurons in a random manner during training for preventing overfitting and promote generalization.
- Flatten Layer: This layer reshapes the pooled feature maps into a one-dimensional vector suitable for the fully connected layers.

- Fully Connected Dense Layer: A fully connected layer with 64 neurons using ReLU activation to learn complex feature interactions. The final output layer with 3 neurons and softmax activation, suitable for multi-class classification (e.g., positive, neutral, negative sentiment).
- Compilation and Training Settings Our model is compiled with Adam optimizer and sparse categorical crossentropy loss, appropriate for multi-class classification with integer-encoded targets. Training is done over ten epochs having a batch size of 1024, and an EarlyStopping callback monitors validation loss with a patience of three and minimum delta of 0.1, stopping early if performance plateaus.

This representation was retained as the sentiment embedding to be fused with numerical data.

*Part 2: Stock Forecasting via LSTM*

The output of the CNN was concatenated with the corresponding day's normalized numerical feature from Yahoo Finance (closing price). The concatenated vectors for each day formed a multivariate time series, which was then fed into the LSTM for forecasting.

The LSTM architecture included:

- Input Layer: Accepts sequences of stock closing prices over a 10-day window, each appended with the corresponding day's FinBERT sentiment score. This results in 11 features per sequence.
- First LSTM Layer: Contains 70 units with tanh activation, allowing the layer to output the full sequence for deeper temporal learning. This layer captures long-range dependencies in the price-sentiment sequences.
- Second LSTM Layer: Comprises 30 units and similarly uses tanh activation with return_sequences=True, refining temporal patterns detected by the first layer.
- Third LSTM Layer: A smaller 10-unit layer with return_sequences=False, compressing sequence-level features into a single output vector for regression.
- Dense Output Layer: A fully connected layer with a single neuron and linear activation function outputs the final predicted stock closing price.
- Compilation and Training Settings: The model is compiled with the Adam optimizer (learning rate = 0.02) and Mean Squared Error (MSE) loss, appropriate for continuous regression tasks. The model was trained over 100 epochs. This architecture was employed in both hybrid models: one leveraging VADER-CNN-based sentiment scores and the other using FinBERT-based sentiment scores, enabling a comparative assessment of sentiment-driven forecasting performance.

Hybrid Model 2: FinBERT-LSTM

*Part 1: Sentiment Processing via FinBERT (No CNN)*

The second hybrid model utilized FinBERT, a transformer-based language model fine-tuned on financial texts, to generate daily sentiment scores directly from headlines. Each day's 10 headlines were input into FinBERT, which assigned sentiment labels (positive, negative, neutral) along with confidence scores.

The scores were aggregated and averaged to yield a single continuous sentiment score per time step, normalized on a scale from -1 to 1. This final score captures the composite financial sentiment conveyed by multiple news items for that time period.

The final sentiment scores were added to the dataset, being utilized as input features for FinBERT-LSTM. This integration allowed the model to learn temporal dependencies not only from stock price movements but also from textual sentiment, offering a solid basis for forecasting.

*Part 2: Forecasting via LSTM (Same Architecture as 5.1)*

Next, the daily sentiment scores from FinBERT were combined with the same numerical feature that the CNN-LSTM model used. For forecasting, the multivariate sequence was put into the identical LSTM architecture. By doing this, it allowed researchers to isolate the differences in results from using different sentiment analysis strategies. The LSTM model used here also had three layers (70, 30 and 10 units) with nonlinear 'tanh' activation and a final dense output layer.

| **Feature** | **VADER - CNN - LSTM** | **FinBERT - LSTM** |
|---|---|---|
| Sentiment Engine | VADER (rule-based) | FinBERT (transformer) |
| Text preprocessing | Lightweight | Tokenization, FinBERT - specific |
| Feature Extraction | CNN | Direct from FinBERT |
| Forecasting Engine | LSTM | LSTM |
| Input Dimensionality | 10 VADER scores | FinBERT - weighted average |
| Sentiment Accuracy | Lower (general purpose) | Higher (financial domain-specific) |

**Table 1.** Comparison of hybrid models

To wrap up the section, various aspects of the models built in the experiment have been described in Table 1. While LSTM model for final forecasting stays identical, the initial components differ from each other to some extent.

## 3. Results and discussion

A more detailed assessment of the two proposed hybrid models has been presented in this part—VADER-CNN-LSTM and FinBERT-LSTM—for forecasting S&P500 closing prices using integrated numerical market data and news sentiment features. Performance was quantified through three key metrics:

1. Mean absolute error is absolute deviation between predicted and actual prices

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{1}$$

Here,

$y_i = actual\ value$

$\hat{y}_i = predicted\ value$

$n = number\ of\ predictions$

2. Mean absolute percentage error is the relative error as a percentage of actual prices.

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{2}$$

3. Accuracy refrects directional correctness:

$$Accuracy = \frac{1}{n-1}\sum_{i=2}^{n} 1[sign(\hat{y}_i - \hat{y}_{i-1}) = sign(y_i - y_{i-1})] \tag{3}$$

*Performance of VADER-CNN-LSTM*

MAE = 44.58 demonstrates that the model's predictions deviated from the real S&P500 closing prices by approximately 44.58 index points on average.

MAPE = 0.0112 shows that, averagely, the model's predictions were 1.12% off from the actual values in percentage terms.

Accuracy = 98.88% reflects a near-perfect directional accuracy, indicating the model was highly effective in predicting the movement direction of the index — a crucial factor for trading strategies.

The model's strong performance is attributed to its dual-stage feature refinement: VADER provided solid foundational sentiment signals using lexicon-based scoring, whereas CNN effectively extracted spatial patterns in news text and filtered out irrelevant noise that does not impact market movement.

| Metric | VADER-CNN-LSTM | Finbert-LSTM | Δ |
|---|---|---|---|
| MAE | 44.58 | 66.79 | 22.21 |
| MAPE | 1.12% | 1.17% | 0.05% |
| Accuracy | 98.88% | 98.32% | -0.56% |

**Table 2.** Summary of the models' relative performance

*Performance of FinBERT-LSTM model*

MAE = 66.79 means the model's predictions deviated from the actual S&P500 closing prices by approximately 66.79 index points on average.

MAPE = 0.0168 (1.17%) shows that averagely, the model's predictions were 1.68% off from the actual values in percentage terms.

Accuracy = 98.32% means the model correctly predicted the direction of market movement 98.32% of the time.

The 22.21% higher MAE (66.79 vs. 44.58) suggests reduced precision compared to VADER-CNN-LSTM. While accuracy remained high (98.32%), the MAPE gap (0.0112 vs. 0.0168) implies marginally weaker performance in volatile regimes.

Table 2 clearly indicates the results of the experiment to highlight the differences between the predictive power of the 2 hybrid models.

## 4. Conclusion

This study examined two hybrid forecasting pipelines—VADER-CNN-LSTM and FinBERT-LSTM—for predicting S&P 500 closing prices using a combination of past prices data and sentiment features extracted from financial news. Both models achieved high predictive performance, with accuracy levels exceeding 98%, underscoring the potential of sentiment-enriched architectures in financial time series forecasting.

VADER-CNN-LSTM demonstrated marginally stronger performance, achieving a Mean Absolute Error of 44.58 and a Mean Absolute Percentage Error of 1.12%, while the FinBERT-LSTM model yielded an MAE of 66.79 and a MAPE of 1.17%. These error levels suggest both models possess high fidelity in capturing short-term price dynamics. The near parity in accuracy (98.88% vs. 98.32%) reinforces the robustness of both approaches, with only minimal variance between them.

Importantly, the results validate the central hypothesis: that incorporating sentiment features into deep learning models can significantly enhance predictive insight into financial markets. The little differences in error rates also show how complexity in models can make them fit new datasets more flexibly. VADER has lexicon-based easiness, which was integrated with CNN's feature extraction capabilities, resulted in lower error margins, whereas FinBERT's transformer-based embeddings still performed well, mainly in handling structured financials. But FinBERT's benefits may not be undertaken in the absence of enough data. The bigger error numbers in the FinBERT-LSTM model show that the quality of a model is impacted by the quantity and structure of the data. This highlights a key point, when the model complexity is aligned with the data scale, sentiment-enhanced forecasting models perform better.

To wrap up, the outcomes indicate that both of the models are able to forecast effectively, with high accuracy. This study backs up the usefulness of combining different types of sentiment signals in predicting stock market movements and provides a base for further improvements including dynamic sentiment weighting or attention-based architectures designed for evolving market trends. Yet, there were some shortcomings. Most news-based reactions to the market are not real-time and the models cannot respond much to unexpected events like geopolitical crises. Studies might gain insights and better capture patterns by using multimodal sentiment analysis - integrating financial reports and instant data from social media along with the news articles.

**Authors' Declaration**

The authors declare no conflict of interests regarding the publication of this article.

## References

1. Fama, E. F. (1970). Efficient Capital Markets: A review of theory and Empirical work. *The Journal of Finance*, *25*(2), 383. https://doi.org/10.2307/2325486
2. Shiller, R. J. (2003). From efficient markets theory to behavioral finance. *The Journal of Economic Perspectives*, *17*(1), 83–104. https://doi.org/10.1257/089533003321164967
3. Wang, M. (2022). Prediction of the Technology Company's Stock Price through the Deep Learning Method. *Open Journal of Modelling and Simulation*, *10*(04), 428–440. https://doi.org/10.4236/ojmsi.2022.104024
4. Fadlalla, A., & Lin, C. (2001). An analysis of the applications of neural networks in finance. *INFORMS Journal on Applied Analytics*, *31*(4), 112–122. https://doi.org/10.1287/inte.31.4.112.9662
5. Tetlock, P. C. (2005). Giving content to investor sentiment: The role of media in the stock market. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.685145
6. Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007
7. Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, *66*(5), 1461–1499. https://doi.org/10.1111/j.1540-6261.2011.01679.x
8. Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news. *ACM Transactions on Office Information Systems*, *27*(2), 1–19. https://doi.org/10.1145/1462198.1462204
9. 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. https://doi.org/10.1111/j.1540-6261.2010.01625.x
10. Noorbakhsh, A. and Shaygani, M. (2024). Deep Learning Approach for Stock Price Prediction: Comparing Single and Hybrid Models. *Advances in Industrial Engineering*, *58*(1), 237-249. https://doi.org/10.22059/aie.2024.372470.1889
11. Chong, E., Han, C., & Park, F. C. (2017). Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems With Applications*, *83*, 187–205. https://doi.org/10.1016/j.eswa.2017.04.030
12. Lee, C., Sehwan, Y., & Jongdae, J. (2007). Neural Network Model vs. SARIMA Model In Forecasting Korean Stock Price Index (KOSPI). *Issues in Information System*, *8*(2), 372–378. URL:https://www.researchgate.net/publication/237820911_Neural_Network_Model_vs_SARIMA_Model_In_Forecasting_Korean_Stock_Price_Index_KOSPI

13. Agrawal, J., Chourasia, D.V., & Mittra, D.A. (2013). State-of-the-Art in Stock Prediction Techniques. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Energy, 2*, 1360-1366. URL: https://www.researchgate.net/publication/287653271_State-of-The-Art_in_stock_prediction_techniques
14. Thawornwong, S., & Enke, D. (2003). The adaptive selection of financial and economic variables for use with artificial neural networks. *Neurocomputing*, *56*, 205–232. https://doi.org/10.1016/j.neucom.2003.05.001
15. Enke, D., & Mehdiyev, N. (2013). Stock market prediction using a combination of stepwise regression analysis, differential evolution-based fuzzy clustering, and a fuzzy inference neural network. *Intelligent Automation & Soft Computing*, *19*(4), 636–648. https://doi.org/10.1080/10798587.2013.839287
16. Thakkar, A., & Chaudhari, K. (2021). A comprehensive survey on deep neural networks for stock market: The need, challenges, and future directions. *Expert Systems With Applications*, *177*, 114800. https://doi.org/10.1016/j.eswa.2021.114800
17. Akita, R., Yoshihara, A., Matsubara, T., & Uehara, K. (2016). Deep learning for stock prediction using numerical and textual information. *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, 1–6. https://doi.org/10.1109/ICIS.2016.7550882
18. Mittal, A., & Goel, A. (2012). *Stock prediction using Twitter sentiment analysis* (CS229 Project). Stanford University. http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf

    Nabipour, M., Nayyeri, P., Jabani, H., Mosavi, A., Salwana, E., & S, S. (2020). Deep learning for stock market prediction. *Entropy*, *22*(8), 840. https://doi.org/10.3390/e22080840
19. Liu, Z., Huang, D., Huang, K., Li, Z., & Zhao, J. (2020, July). FinBERT: A pre-trained financial language representation model for financial text mining. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)* (pp. 4513–4519). International Joint Conferences on Artificial Intelligence Organization. https://doi.org/10.24963/ijcai.2020/622
20. Hu, Z., Liu, W., Bian, J., Liu, X., & Liu, T.-Y. (2018). Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (WSDM '18)*, Association for Computing Machinery, New York, NY, USA, 261–269. https://doi.org/10.1145/3159652.3159690
21. Liu, H. (2018). Leveraging Financial News for Stock Trend Prediction with Attention-Based Recurrent Neural Network. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1811.06173
22. Lee, J., Kim, R., Koh, Y., & Kang, J. (2019). Global stock market prediction based on stock chart images using deep Q-Network. *IEEE Access*, *7*, 167260–167277. https://doi.org/10.1109/access.2019.2953542
23. Araci, D. (2019). FinBERT: Financial Sentiment Analysis with Pre-trained Language Models. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1908.10063
24. Théate, T., & Ernst, D. (2021). An application of deep reinforcement learning to algorithmic trading. *Expert Systems With Applications*, *173*, 114632. https://doi.org/10.1016/j.eswa.2021.114632

25. Pricope, T. (2021). Deep Reinforcement Learning in Quantitative Algorithmic Trading: a review. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2106.00123
26. Zhu, P., Li, Y., Hu, Y., Liu, Q., Cheng, D., & Liang, Y. (2024). LSR-IGRU: Stock Trend Prediction based on Long Short-Term Relationships and Improved GRU. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2409.08282
27. Zhang, Chi. (2025). Dynamic Asset Pricing: Integrating FinBERT-Based Sentiment Quantification with the Fama--French Five-Factor Model. *arXiv (Cornell University)*. https://doi.org/10.48550/arXiv.2505.01432
28. Shobayo, O., Adeyemi-Longe, S., Popoola, O., & Ogunleye, B. (2024). Innovative sentiment analysis and prediction of stock price using FINBERT, GPT-4 and Logistic Regression: a Data-Driven Approach. *Big Data and Cognitive Computing*, *8*(11), 143. https://doi.org/10.3390/bdcc8110143
29. Halder, S. (2022). FinBERT-LSTM: Deep Learning based stock price prediction using News Sentiment Analysis. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2211.07392
30. Gupta, H., & Jaiswal, A. (2024). A study on stock forecasting using deep learning and statistical models. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2402.06689
31. Jing, N., Wu, Z., & Wang, H. (2021). A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Systems With Applications*, *178*, 115019. https://doi.org/10.1016/j.eswa.2021.115019
32. Pooja, Nagalakshmi Vallabhaneni, Harshini S R, Jayaprakash. (n.d.). Stock market prediction using deep learning. School of Information Technology and Engineering. Retrieved from: https://www.scribd.com/document/751240131/220746-Stock-Market-Prediction-Using-Deep-Learning
33. Wu, Y., Sun, M., Zheng, H., Hu, J., Liang, Y., & Lin, Z. (2024). Integrative analysis of financial market sentiment using CNN and GRU for risk prediction and alert systems. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2412.10199
34. Patil, S. S. (2023). Convolutional Neural Networks for Text Classification: A Comprehensive analysis. *International Journal for Research in Applied Science and Engineering Technology*, *11*(5), 6039–6047. https://doi.org/10.22214/ijraset.2023.52889