



Stock Price Trend Prediction Using LSTM and Sentiment Analysis on News Headlines

Jung-Bin Li¹, Szu-Yin Lin², Fang-Yie Leu^{3,4}(✉), and Yen-Chu Chu¹

¹ Department of Statistics and Information Science, Fu Jen Catholic University, New Taipei City, Taiwan
071635@mail.fju.edu.tw

² Department of Computer Science and Information Engineering, National Ilan University, Yilan, Taiwan

³ Department of Computer Science, Tunghai University, Taichung, Taiwan
leufy@thu.edu.tw

⁴ Emergency Response Management Center, Ming Chung University, Taoyuan, Taiwan

Abstract. To simulate the trading behavior of investors in the stock market, this study adopts parameters including technical, fundamental, and chip to build a LSTM model, and also observes the ability of news sentiment to predict stock prices. Influential stocks such as TSMC, Fulgent Sun, and HTC are chosen as the target of our experiment. Four common natural language processing packages are used to label news sentiment. Then the combined sentiment labels along with the LSTM model are used for backtesting. The results of the study found that FinBERT's ability to predict the price trend outperforms other methods, with an accuracy of 41.6%. In addition, combining news sentiment labels with the LSTM model generally leads to better outcome than using either the news label or the LSTM model alone. However, in certain extreme cases, traditional technical indicators or even buy-and-hold strategy have better performances.

1 Introduction

Due to the impact of economic and pandemic factors in recent years, coupled with the background factors such as the opening of the Taiwan stock market for intraday fractional trading and the lowering of trading thresholds, more people have been attracted. These novice investors tend to seek simple and efficient ways to make investment decisions. Statistics revealed from the Taiwan Stock Exchange show that by the end of 2021, the total number of accounts opened has exceeded 12 million, and the number of newly opened accounts has reached 770,000, indicating that the general public's willingness to invest in the stock market has increased considerably.

In the past, some studies have used linear regression [1] or Support Vector Machine [2], k-nearest [3] and other models to predict stock prices, and the random walk hypothesis [4], and some also have tried the ARIMA (Autoregressive Integrated Moving Average model) [5]. RNN (Recurrent Neural Network) is one of the common methods in natural language processing. Since stock prices and news posts are convincingly related data, some studies have tried to use LSTM (Long Short-Term Memory) to predict stock

prices [6]. Similar studies in the past mostly focused on English reports and tended to use technical parameters. This research hopes to use more comprehensive information to improve the accuracy of the LSTM model for stock price prediction to assist investors.

2 Related Works

2.1 Long Short-Term Memory (LSTM)

Hochreiter and Schmidhuber proposed LSTM in 1997 [7]. The core concept is that the cell state trades off the incoming information from the hidden layer, so that the problem of dependence between the weights can be solved. Many derivative models of LSTM have been proposed. Graves and Schmidhuber raised Bi-LSTM (Bidirectional LSTM) to analyze phonemes [8]. Greff, Srivastava, Koutnik, Steunebrink, and Schmidhuber tried to change the structure and various parameters of LSTM, and the results showed that the change of its structure could not significantly improve the learning ability of the original model, and the parameters used in fine-tuning were closely related to the training data [9]. Graves, Jaitly, and Mohamed used Bi-LSTM for speech recognition [10], and Shi, Chen, Wang, Yeung, Wong, and Woo applied LSTM to rainfall prediction [11]. Liang, Shen, Feng, Lin, and Yan contributed to the improvement of LSTM in image processing [12].

There are studies applying deep learning to stock price prediction. Liu, Liao, and Ding believe that in the stock price prediction task, the accuracy of multi-layer LSTM will be higher than that of single-layer LSTM, but the cost will also increase [13]. Ojo, Owolawi, Mphahlele, and Adisa used LSTM to predict stock prices and suggested adding factors such as international news and policies as directions for improvement [14]. Bathla [15] compares the stock price prediction ability between LSTM and SVM models, and the results show that LSTM outperforms SVM. Guo added news sentiment labels as input parameters to the LSTM model, and its prediction performance was significantly improved compared to purely using stock daily trading information as input [16]. Kavinnilaa, Hemalatha, Jacob, and Dhanalakshmi use LSTM to predict stock prices, and point out that stock prices are affected by news in the market, and propose that significant information events can be set as parameters to improve model accuracy [17].

2.2 Natural Language Processing (NLP)

Dervlin, Chang, Lee, and Toutanova proposed BERT (Bidirectional Encoder Representations from Transformers) in 2018, which mainly uses the Encoder part of Transformer to generate language models. Since the training of BERT requires a lot of computing resources, most studies use pre-trained models that have been trained for feature extraction as input parameters for fine-tuning training [18]. Araci screened keywords in the financial field and established a TRC2-financial dataset, which contains more than 29,000,000 words and nearly 400,000 sentences, making FinBERT a pre-trained BERT model in the financial field [19]. Therefore, this study applies FinBERT and other NLP methods with the attention to reach good results.

3 Proposed Model and Mechanisms

3.1 Model Framework

In this study, the parameters were first imported from the relevant databases, and the data was preprocessed. Then we use the LSTM model to train and compare the differences among parameter combinations. Our LSTM model takes the daily closing price of a stock as the dependent variable, and the independent variables are retrieved from different perspectives. In addition, we observe whether the sentiment of news headlines is helpful for stock price prediction, and then use the output of the model as the basis for stock operation to analyze and compare the backtest results. Model framework is illustrated in Fig. 1.

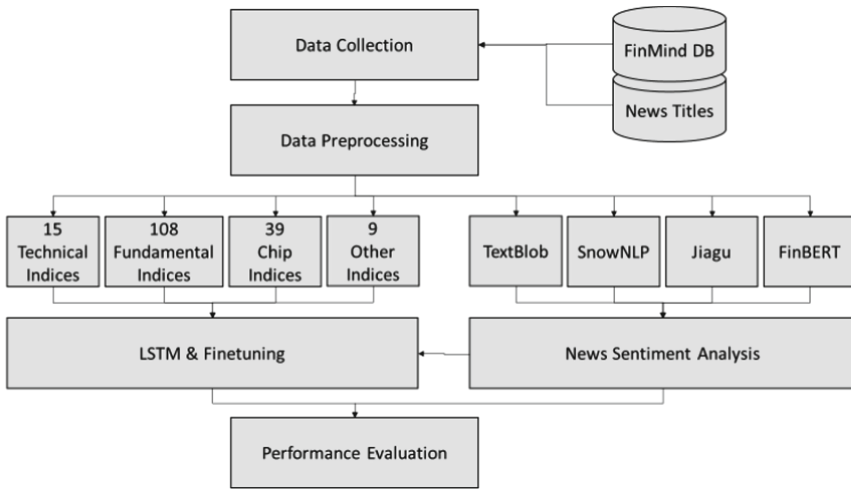


Fig.1. Model framework.

3.2 Data Source

This study uses data from the FinMind database to obtain 171 variables from technical, fundamental, and chip perspectives; news sources are obtained from 23 online news databases. We use these variables to form the reference factors for investment decisions.

Technical. Using the stock price daily transaction information (8), individual stock PER and PBR data (3), the day’s offset transaction target and trading volume value (3), and the weighted stock price index, a total of 15 variables are used. The numbers in brackets represent the number of variables adopted. The daily transaction information table of stock prices includes transaction volume, transaction amount, opening price, highest price, lowest price, closing price, the difference from the previous day’s closing price, and the number of transactions. The PER and PBR data of individual stocks include cash

yield, price-to-earnings ratio, and price-to-book value ratio. The target and volume value of the offset transaction on the day include the transaction volume, the buying amount, and the selling amount.

Fundamental. A total of 108 variables are obtained from the consolidated income statement (25), balance sheet (44), cash flow statement (33), dividend policy and ex-rights and ex-dividend results (4), and monthly income statement (2).

Chip. A total of 39 variables were obtained from the margin financing and securities lending table (11), legal person trading (5), corporate trading in the Taiwan market as a whole (6), foreign shareholding (2), and shareholders' shareholding classification table (15).

News. Stock-related Mandarin news headlines from 23 sources including Anue, EBC, ETtoday, HiNet News Community, INSIDE, MoneyDJ, NOWnews, UDN, YahooNews, ChinaTimes, BusinessToday, TaiwanNews, AppleDaily, BusinessWeekly, Commonwealth, CommercialTimes, NewTalk, MoneyWeekly, TechNews, ManagerToday, LibertyTimes, Sina, and GlobalViews are collected.

Others. Nine variables including the exchange rate of Taiwan dollar against the US dollar, the US Federal Reserve interest rate, the US one-month Treasury bond yield, the price of gold, the price of crude oil, the CNN Fear and Greed Index, the US stock TSM ADR, the Dow Jones Index, and the Philadelphia Semiconductor Index are adopted.

3.3 Data Processing

Preprocessing. Because news reports are usually published outside the stock trading day, we fill in its 30-day moving average if the stock information does not exist during the unopened period. Quarterly data will be repeatedly filled downwards until the next season's release date.

News Sentiment Labels. This research uses TSMC (2330), which is the largest market capitalization in Taiwan stock market, as the main object. Benefiting from TSMC's trading volume, its related news is ample. The research retrieves news from July 28, 2020 to May 1, 2022, with a total of 10,381 records. Cleaning works including removing duplicate titles, stock codes, news webpage links, data sources and the data source part in the title are done.

This study applies different methods to label news headlines with sentiment. The first method is to use TextBlob to get news headline sentiment scores. TextBlob is a text processing tool based on two python packages, NLTK and Pattern. The TextBlob output format is as follows:

Sentiment (polarity = 0.3916, subjectivity = 0.4357)

The polarity score ranges from -1 to 1 . The closer to -1 , the more pessimistic the sentiment is; the closer to 1 , the more optimistic the sentiment. Subjectivity ranges from 0 to 1 , the closer it is to 0 , the more objective it is.

The second is SnowNLP, which is similar to TextBlob. The training of sentiment analysis tasks uses the Bayesian model and its self-equipped dictionary. Its score represents the probability of being a positive sentiment, ranging from 0 to 1. The closer the value is to 1, the more positive it is.

The third is Jiagu. It also supports the function of customizing the new dictionary, but is based on BiLSTM training. After Jiagu word segmentation, the idea of bag of words is used, and the NTUSD emotional vocabulary dictionary from the Taiwan Natural Language Processing Laboratory is used for classification. The negative words are given a score of -1, and the positive words are given a score of 1. For neutral words, a score of 0 is given.

The fourth is to use FinBERT to get news headline sentiment scores. FinBERT is a pre-trained model for analyzing financial text sentiment. Its output format is as follows: { 'label': 'Negative', 'score': 0.9966173768043518 }

The label has three possible sentiments: negative, positive, and neutral. Negative words are classified to negative and given a score of -1; positive words are classified to positive and given a score of 1; and neutral words are classified to neutral. With score 0. The score represents the intensity of sentiment. This study adopts the label column only.

3.4 Data Analysis

In this study, TSMC (2330) is chosen to evaluate our model architecture. In order to observe the model performance, we additionally take Fulgent Sun-KY (9802) and HTC (2498) as supplement for the analysis. This research collected a total of 603 data tuples from July 28, 2020 to May 1, 2022. The input fields cover technical, fundamental, chip, news, and other variables, with a total of 171 variables. The data is divided into training data set and verification data set with the ratio 9 to 1.

Subsequently, we use the TSMC news data to compare the prediction ability of each news sentiment labeling method, and then use the most effective labeling method to train the LSTM model and back-test for comparison.

4 Model Performances and Findings

4.1 LSTM Outcome

This study trains and compares models using different input and output lengths, including the following combinations: 1-to-1, 1-to-5, 5-to-1, 5-to-5, 30-to-1, and 30-to-5. One-to-one features the model of using data of one day in the past to predict the price trend of one day in the future. To simplify the description, subsequent references to the model will denote the three-layer 1-to-1 LSTM as LSTM3 (1, 1).

All LSTMs use Adam (Adaptive Moment Estimation) as the Optimizer, and the dropout ratio in the architecture is set to 0.2. We add Early Stopping to prevent overfitting, and training will be stopped when the loss has not been improved after 10 epochs. The comparison between the models is based on RMSE, and the results are shown in Table 1.

Table 1. Comparison of LSTM models.

Model	RMSE	Model	RMSE
LSTM1 (1, 1)	33.2822	LSTM1 (5, 5)	64.3136
LSTM3 (1, 1)	20.2428	LSTM3 (5, 5)	50.9448
LSTM5 (1, 1)	20.3868	LSTM5 (5, 5)	52.1754
LSTM1 (1, 5)	131.4313	LSTM1 (30, 1)	35.6099
LSTM3 (1, 5)	32.7369	LSTM3 (30, 1)	27.8210
LSTM5 (1, 5)	40.3517	LSTM5 (30, 1)	28.6284
LSTM1 (5, 1)	42.8656	LSTM1 (30, 5)	49.8965
LSTM3 (5, 1)	48.6941	LSTM3 (30, 5)	38.8199
LSTM5 (5, 1)	51.9153	LSTM5 (30, 5)	34.3075

In this part of the experiment, the following findings were observed:

1. Increasing the number of layers of the LSTM model did not always improve the performance of the mode. There is little difference between the three-layer and five-layer LSTM. Considering the computational efficiency, subsequent experiments will be carried out with three-layer LSTM.
2. The multi-to-1 models are not as good as using the 1-to-1 model. Among our experimental models, the LSTM3 (1, 1) model has the best predictive ability (RMSE = 20.2428).
3. The results of multi-to-multi and less-to-more models are in average worse than other combinations, and the maximum RMSE is from LSTM1 (1, 5).

4.2 News Sentiment

In the news sentiment analysis module, this research uses four sentiment labels, namely FinBERT, Jiagu, TextBlob, and SnowNLP, to classify a total of 10,385 TSMC news headlines from July 28, 2020 to May 1, 2022. In order to understand the distribution of the data, the results of each label are further classified into interval scales. FinBERT uses its output directly. Jiagu is set neutral when score 0, positive when the score is greater than 0, and negative when less than 0. TextBlob is set neutral when score equals to 0, positive when greater than 0, and negative when less than 0. SnowNLP is neutral when scores are in the range (0.4, 0.6), positive when greater than 0.6, and negative when less than 0.4. The distribution is shown in Table 2.

All labeling methods except SnowNLP have more neutral labels. SnowNLP has more negative labels, and it may be attributed to its dictionary which is majorly based on e-commerce. In the first half of the study's training period, TSMC has a general rise trend. Hence the news media reported mostly good news, and the distribution of news labels fits the actual situation.

Table 3 shows the overall accuracy of news labels. FinBERT owns the highest accuracy, and SnowNLP is the lowest. In terms of precision, again FinBERT is slightly

Table 2. Distribution of news labels.

Method	FinBert	Jiagu	TextBlob	SnowNLP
Positive	2555	3394	4170	2085
Neutral	5981	5134	4463	829
Negative	1849	1857	1752	7471

higher than other methods in neutral and negative news classification. However, Jiagu is slightly higher than FinBERT in the matter of positive news. In terms of recall, Jiagu is the highest for positive news, FinBERT for neutral, and SnowNLP for negative. Jiagu may be benefited by its customizable positive dictionary. The number of negative news classified by SnowNLP is significantly higher than that of other methods, which causes its higher recall.

Table 3. Comparison of news sentiment labels.

Method		FinBert	Jiagu	TextBlob	SnowNLP
Accuracy		0.416177	0.379104	0.36564	0.260664
Precision	Positive	0.260281	0.280022	0.220320	0.243073
	Neutral	0.534024	0.509349	0.532153	0.496984
	Negative	0.252838	0.236299	0.245803	0.229736
Recall	Positive	0.192800	0.208000	0.154400	0.012630
	Neutral	0.590825	0.483721	0.439327	0.093049
	Negative	0.260589	0.323518	0.413473	0.479391

4.3 Model Backtest

This study collects the price data of TSMC from March 3, 2022 to May 1, 2022. The initial holding capital is set at 10 million NTD. Other assumptions include the securities transaction tax as 0.3% of the transaction amount, and the brokerage fee is 0.1425% of the transaction amount. We use the 1-to-1 3-layer LSTM model. When the predicted value is greater than the actual stock price, it is seen as a buy signal. When the predicted value is less than the actual stock price, it is set as a sell signal. When the predicted value is equal to the actual stock price, we hold without trading.

Then the FinBERT news sentiment label is used as an operation reference. We buy when the sentiment label is positive; we sell when it is negative; and we hold when it is neutral. Additionally, we make decisions based on the FinBERT news sentiment label combined with the result of the LSTM model. It triggers a buy signal when the sentiment label is positive and the LSTM prediction result is greater than the actual stock price.

Sell signals ring when the sentiment label is negative and the LSTM prediction result is lower than the actual stock price.

The backtest results of the above strategies are summarized in Table 4. All models end up with negative earnings. The reason is that TSMC's price trend is downward during the testing period, indicating that the overall trend of the investment target has a direct impact on the final profit, and its timing of buying is equally important. However, for the three models of LSTM, news sentiment label, and LSTM plus sentiment label, the LSTM plus sentiment label model is better than the others.

Table 4. Backtest results of TSMC.

Model	LSTM	News label	LSTM + News label	Buy & hold
Average profit (NTD)	-81754	-76106	-44214	-22079
Max loss (NTD)	724128	283924	258855	62693
Average profit rate (%)	-0.82	-0.76	-0.44	-0.22
Net profit margin (%)	-4.36	-2.03	-1.74	-0.46
Internal rate of return (%)	-23.75	-11.73	-10.13	-2.77
Sharpe index	-1.34	-2.66	-2.24	-1.81

To further validate the performance of our model, Fulgent Sun (9802), which had an upward stock price trend, and HTC (2498), which fluctuated frequently, were used as comparisons. The results are shown in Table 5. For the backtest of 9802, the performance ranking is news, LSTM plus news, and LSTM. The result is different from that of TSMC. HTC's backtest results are similar to those of TSMC, and LSTM and news sentiment labels cannot capture the starting point of the stock price. But what is certain from this

Table 5. Back test results of 9802 and 2498.

StockID	9802			2498		
	LSTM	News label	LSTM + News label	LSTM	News label	LSTM + News label
Average profit (NTD)	78912	260858	154183	-424085	-39075	-23263
Max loss (NTD)	68120	67430	54173	157342	167224	97219
Average profit rate (%)	0.79	2.61	1.54	-0.42	-0.39	-0.23
Net profit margin (%)	6.44	11.82	8.2	-1.43	-1.35	-0.8
Internal rate of return (%)	37.22	76.18	49.11	-7.14	-6.75	-4.05
Sharpe index	4.87	6.18	5.69	-2.9	-2.42	-2.5

performance comparison is that adding news sentiment labels does have a positive effect on stock price prediction. This finding stands in the backtest of all three stocks of the study.

5 Conclusion

The first part of this study tried to use LSTM models with different layers to predict the stock price of TSMC (2330), and found that the 1-to-1 3-layer model outperformed other combinations, and the 1-to-1 5-layer model had a slightly higher RMSE than 3-layer model. The 30-to-1 model, on the other hand, is not as predictive as the 1-to-1 model. It is speculated that some parameters generate noise on the prediction results during learning, which has an impact as the sampling time lengthens. In the comparison of different news sentiment labels, including the prediction of stock price fluctuations, the accuracy rate of FinBERT is the highest, which is 41.6%. This is not good enough in our experiment, so we have to give investors some reminder that it is not a reliable reference to make investment decisions solely based on news.

For 2330, all three models in the backtesting stage ended up with negative profits, because its stock price trended down significantly during the testing period. Compared with the LSTM model, the LSTM plus news labels model has more restrictions on buying and selling conditions, and it is less likely to trigger buying and selling behavior and reduce losses. The LSTM and LSTM plus news labels models continued to buy during the period, and the profit situation was similar to the buy-and-hold method. The unrealized loss of the news model is smaller than that of LSTM and LSTM plus news labels, but the realized profit and loss of LSTM and LSTM plus news labels is higher than that of news. Regarding to the average profit result, LSTM plus news labels is better than LSTM and news strategy.

Compared with 2330, the results of Fulgent Sun (9802) and HTC (2498) vary in model performance ranking. However, it can be observed that the addition of the news sentiment label model does help profit. This outcome inspires us to further extend and fine-tune our model.

References

1. Bhuriya, D., Kaushal, G., Sharma, A., and Singh, U.: Stock market predication using a linear regression. In: 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), pp. 510–513 (2017)
2. Lai, L.K.C., and Liu, J.N.K.: Stock forecasting using support vector machine. In: 2010 International Conference on Machine Learning and Cybernetics, pp. 1607–1614 (2010)
3. Taunk, K., De, S., Verma, S., and Swetapadma, A.: A brief review of nearest neighbor algorithm for learning and classification. In: 2019 International Conference on Intelligent Computing and Control Systems (ICCS), pp. 1255–1260 (2019)
4. Fama, E.F.: Random walks in stock market prices. *Financ. Anal. J.* **21**(5), 55–59 (1965)
5. Ariyo, A.A., Adewumi, A.O., and Ayo, C.K.: Stock price prediction using the ARIMA model. In: 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, pp. 106–112 (2014)

6. Chen, K., Zhou, Y., and Dai, F.: A LSTM-based method for stock returns prediction: a case study of China stock market. In: 2015 IEEE International Conference on Big Data (Big Data), pp. 2823–2824 (2015)
7. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997)
8. Graves, A., Schmidhuber, J.: Framewise phoneme classification with bidirectional LSTM networks. In: Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN), Montreal, QC, Canada, Vol. 4, pp. 2047–2052 (2005)
9. Greff, K., Srivastava, R.K., Koutník, J., Steunebrink, B.R., Schmidhuber, J.: LSTM: A Search Space Odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **28**(10), 2222–2232 (2017)
10. Graves, A., Jaitly, N., and Mohamed, A.: Hybrid speech recognition with deep bidirectional LSTM. In: 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, pp. 273–278 (2013)
11. Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K. and Woo, W.C.: Convolutional LSTM Network: a machine learning approach for precipitation nowcasting. In: Proceedings of the 28th International Conference on Neural Information Processing Systems (NIPS 2015), vol. 1, 802–810 (2015)
12. Liang, X., Shen, X., Feng, J., Lin, L., Yan, S.: Semantic object parsing with graph LSTM. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9905, pp. 125–143. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_8
13. Liu, S., Liao, G. and Ding, Y.: Stock transaction prediction modeling and analysis based on LSTM. In: 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), pp. 2787–2790 (2018)
14. Ojo, S.O., Owolawi, P.A., Mphahlele, M., Adisa, J.A.: Stock market behaviour prediction using stacked LSTM networks. 2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC), pp. 1–5 (2019)
15. Bathla, G.: stock price prediction using LSTM and SVR. In: 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC), pp. 211–214 (2020)
16. Guo, Y.: Stock price prediction based on LSTM neural network: the effectiveness of news sentiment analysis. In: 2020 2nd International. Conference on Economic Management and Model Engineering (ICEMME), pp. 1018–1024 (2020)
17. Kavinnilaa, J., Hemalatha, E., Jacob, M.S., Dhanalakshmi, R.: Stock price prediction based on LSTM deep learning model. In: 2021 International Conference on System, Computation, Automation and Networking (ICSCAN), pp. 1–4 (2021)
18. Dervlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint [arXiv:1810.04805](https://arxiv.org/abs/1810.04805) (2018)
19. Araci, D.: FinBERT: financial sentiment analysis with pre-trained language models. In: The International Conference on Learning Representations (ICLR) arXiv: 1908.10063v1 cs.CL (2019)