Proceedings of the 3rd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/2025.30672

Stock Price Prediction Report

Shangyuan Liu[1†], Weijian Chen[2†], Lewei Tian[3*†], Junyi Zheng[4†]

[1]*College of Letters and Science, University of Wisconsin–Madison, Madison, USA*
[2]*College of Liberal Arts & Sciences, University of Illinois Urbana-Champaign, Champaign, USA*
[3]*Faculty of Science, Simon Fraser University, Burnaby, Canada*
[4]*Glasgow College, University of Electronic Science and Technology of China, Chengdu, China*
**Corresponding Author. Email: lta61@sfu.ca*
[†]*These authors contributed equally to this work and should be considered as co-first author.*

Abstract. Stock price prediction is a critical aspect of financial markets, attracting the attention of investors, analysts, and researchers. Accurate forecasting of stock prices can lead to significant economic gains, but due to the complexity of the stock market behavior, accurately predicting the stock price is very challenging. Our approach is to find a relatively stable model to predict the stock price. From our early research, we found that among all the models others developed, Long-Short Term Memory (LSTM) based models might be the most efficient models in most circumstances. While only the LSTM model itself can not provide a valid prediction, we tried to find the combination of the LSTM model and other factors.

Keywords: Stock price prediction, GA-LSTM, FinBERT, LSTM, Sentiment analysis.

1. Introduction

Long-Short Term Memory is a type of RNN architecture designed to effectively model sequences and time-series data. The key components of LSTM are cell states which are the memory components that carry information over multiple time steps and 3 related gates - forget gate, input gate, and output gate. These gates regulate the flow of information, deciding which data should be retained, forgotten, or output at each step of the sequence. This gating mechanism helps the LSTM model solve the short-term memory problem of standard RNNs, and make them functional at capturing data that occur over long time horizons.

The ability to learn from the historical data and make predictions to the future makes the LSTM model ideal for the stock price prediction. In this context, the LSTM can process historical stock data, identify the patterns, and generate forecasts. We will introduce various LSTM models to predict stock price in this report.

DOI: 10.54254/2755-2721/2025.30672

2. GA-LSTM

2.1. LSTM and challenges

Long short-term memory (LSTM) is a widely recognized and effective method for time series prediction, especially in scenarios such as stock market forecasting. However, for the model to perform well, parameters including the number of tiers, units per tier, learning rate, dropout rate must be fine-tuned because the effectiveness of LSTM largely depends on the choice of optimal hyperparameters.

Manually selecting these hyperparameters is not only time consuming, but may not result in an optimal configuration, which in turn affects model performance. Therefore, a method that automates and optimizes hyperparameter selection can effectively improve LSTM.

2.2. GA-LSTM

One of the above methods to improve LSTM defects by optimizing parameters is GA (Genetic Algorithm). Genetic algorithm is a heuristic optimization technique that mimics the process of natural selection, making it particularly useful in successive generations of hyperparameter evolution: it can find more efficient optimal configuration searches faster and better than manual tuning or grid search methods.

GA-LSTM applies this algorithm to the tuning of LSTM and evaluates each set of hyperparameters according to the fitness function (usually the mean square error (MSE)). Through this evaluation, the parameters can be further iteratively optimized to reduce the error in the prediction, to obtain a more accurate and efficient model and a higher prediction accuracy [1].

2.3. Validation and results

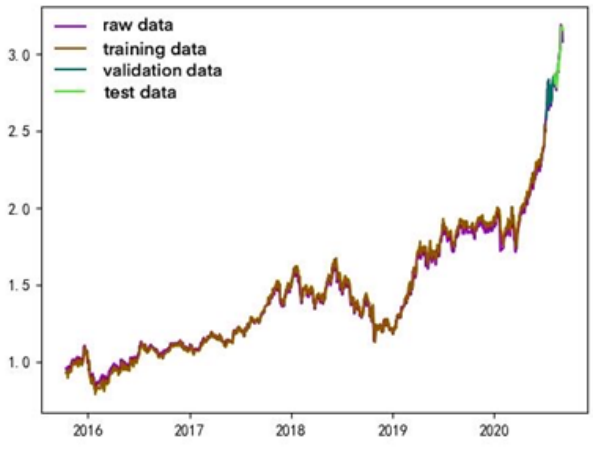

Figure 1. The result of predictions

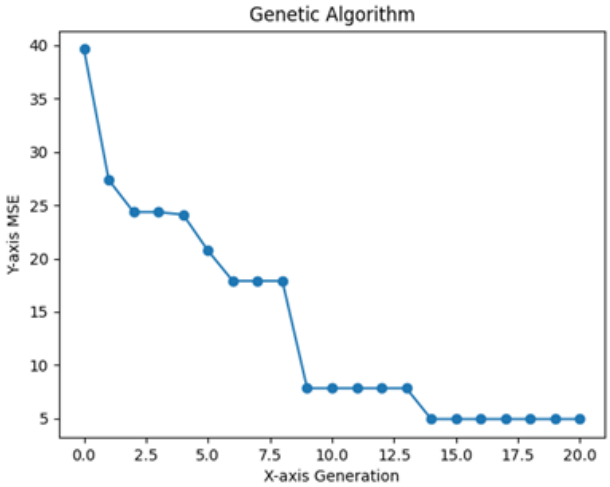

Figure 2. The decrease of MSE [2]

To evaluate the effectiveness of GA-LSTM, we applied the model to the dataset and monitored its performance. As shown in Figure 1, GA-LSTM shows excellent prediction accuracy. One of the most notable observations is the continuous reduction of MSE as genetic algorithms continue to iterate. As shown in Figure 2, starting with a high initial MSE, the error of the model continues to decrease, and after 20 generations, the MSE drops to 4.94.

DOI: 10.54254/2755-2721/2025.30672

These results affirm the ability of GA-LSTM to refine its hyperparameters iteratively and to improve the model's predictive power, which makes it a robust tool for time-series forecasting.

2.4. Limitations of GA-LSTM and discussion

Despite GA-LSTM's apparent improvement in forecasting accuracy, it has a fundamental limitation in its forecasting mechanism: it cannot account for real-world factors that affect stock market trends, such as market sentiment or breaking news events. GA-LSTM focuses only on historical data patterns, which limits its adaptability to sudden, unquantifiable changes in the market. This issue seriously affects the usability and practical value of the model.

To overcome this shortcoming, we shifted our focus to more comprehensive predictive models that integrate real-world influences, such as market sentiment analysis. These models can not only rely on historical data, but also consider qualitative and external factors which may affect the forecast, thus providing a "true" predictive effect.

3. Fin-BERT sentiment analysis

3.1. Introduction

We all know that the stock market can change rapidly. And in the stock market, various factors such as economic updates, news events and investor sentiment can affect stock prices. To address the impact of these factors, we use sentiment analysis as a tool to understand market trends by measuring the tone of news articles that may influence investor behavior, and thus perform predictive analysis. In this project, we use the FinBERT model to study how news headlines affect the stock prices of companies in the S&P 500 index. Our goal is to explore the relationship between financial sentiment and stock price volatility. Along the way, we will also discuss the limitations of FinBERT's model and discuss how the use of more advanced techniques (long short-term memory (LSTM) networks) can better improve prediction accuracy.

3.2. Dataset overview

The data sets we used in our analysis for this project included daily trading data for Apple Inc. (AAPL), and the content of these data sets included important information such as date, ticker symbol, opening price, closing price, trading volume, and percentage of sentiment (positive, negative, and neutral) extracted from the headlines. In these data, the "price change" is represented by binary values: -1 indicates a decline in the share price, while 1 indicates an increase in the share price. We evaluate this data to understand how day-to-day sentiment affects stock performance.

The sentiment column, meanwhile, divided daily news headlines into positive, negative and neutral percentages. Positive sentiment indicates optimistic news about stock market prices, while negative sentiment reflects pessimistic news about stock market prices, and such negative news may lead to negative reactions from investors.

DOI: 10.54254/2755-2721/2025.30672

date	stock	Open	Close	Volume	Positive	Negative	Neutral	Price_change
2020-03-09	AAPL	65.937500	66.542503	286744800	0.046127	0.411465	0.542409	-1
2020-03-10	AAPL	69.285004	71.334999	285290000	0.070845	0.449025	0.480130	1
2020-03-11	AAPL	69.347504	68.857498	255598800	0.190995	0.453761	0.355244	-1
2020-03-12	AAPL	63.985001	62.057499	418474000	0.204221	0.447518	0.348261	-1
2020-03-13	AAPL	66.222504	69.492500	370732000	0.315863	0.218127	0.466010	1

Figure 3. Stock deta example [3]

Figure 4. The relationship between emotion scores [3]

3.3. Analysis

To more intuitively show the relationship between sentiment scores (positive and negative) and stock price movements, we have generated a scatter plot. In the chart, we can see that the X-axis represents positive sentiment and the Y-axis represents negative sentiment. The colors represent different degrees of price change, ranging from -1 (maximum decline) to 1 (maximum increase), and the size of the dots corresponds to the percentage of neutral sentiment.

Our analysis reveals a noteworthy trend: there seems to be a strong correlation between high negative sentiment (greater than 0.5) and price declines, whereas when positive sentiment is moderate, price increases or declines may occur more randomly. This suggests that negative sentiment drives stock price changes more than positive sentiment, and that the behavior of such price changes is consistent with the reality that investors tend to react more strongly to bad news than to good news.

At the same time, we can see that positive sentiment itself seems to be weakly correlated with stock price increases, and the distribution of stock price changes along the positive sentiment axis is

DOI: 10.54254/2755-2721/2025.30672

very diffuse, which is aptly illustrated by this diffuse distribution. This suggests that positive reports do not guarantee an increase in investor confidence or stock performance [4,5].

3.4. Conclusion

Sentiment analysis of financial news headlines using FinBERT can provide valuable insights into stock market performance, particularly in identifying the impact of negative sentiment on price declines. However, the model has strong predictive power only for sentiment analysis, and the model's ability may be lacking for stock prices. Therefore, we should also adopt advanced machine learning techniques (e.g., LSTM) and integrate a variety of financial indicators, which in combination with the LSTM model can provide more comprehensive and accurate predictions of stock price trends.

4. Stock-price-prediction-using-LSTM-and-sentiment-analysis

4.1. Introduction

This model combines sentiment analysis with LSTM (Long Short-Term Memory) for predicting stock prices. First, it analyzes financial news or related posts to explore market sentiment. Then, it incorporates historical stock data as input for the LSTM model to perform analysis. LSTM is a type of recurrent neural network, consisting of multiple layers, which is well-suited for long and volatile time series data like stock prices.

4.2. Datasets

For sentiment analysis, BERT is first used to analyze financial news. VADER is also used to assist in sentiment analysis, especially for short texts that appear in news or posts. The closing price of Yahoo Finance is used as the main input data for this model. In this process, a sliding window is used as a processing method, which predicts future stock prices based on past stock prices. Through this, we can clearly see the time changes of stock price fluctuations.

4.3. Process

The model was first tested on Apple's stock, which used a combination of historical data prices and sentiment analysis to predict future stock market prices. At the same time, predictions were made on Nvidia's stock to test its generalization. In order to test the stability of the model in different markets, the stock of Metro (a food company) and Globalstar (a satellite communications company) was used as input to test it. Finally, the test showed the adaptability of the model in different markets and industries.

4.4. Results

The combination of BERT and VANDER can efficiently obtain more accurate stock market sentiment analysis. At the same time, LSTM captures the fluctuations in past stock price time series. By combining these two, the model effectively captures the patterns of market sentiment and price changes, significantly improving the accuracy of stock price predictions, rather than simply identifying patterns from past data. This method also shows good stability and generalizability.

DOI: 10.54254/2755-2721/2025.30672

4.5. Conclusion

Unlike general models, sentiment analysis-based models use textual data for analysis. By interpreting these sentiments, the model can infer potential stock price movements, particularly in the short term, where market sentiment can change rapidly following major events, as social media updates and news evolve in real time. However, these models are highly sensitive to the quality of textual data; inaccurate or misleading information can distort predictions. Moreover, stock market news is not always perfectly aligned with actual market movements, leading to occasional prediction errors. Therefore, under the right conditions for sentiment analysis, combining it with other modeling techniques offers improved results, balancing short-term sentiment with long-term data trends.

5. Summary

First, we were inspired by reading the paper Deep Learning for Stock Market Prediction. Then, past few weeks, we read many papers about using genetic algorithms (GA), recurrent neural networks (RNN), and long short-term memory networks (LSTM), etc to predict stock prices. We also read the article on two model combinations to predict the stock price, GA-LSTM. Compared to the previous single models, the stock price prediction accuracy of GA-LSTM is better. However, GA-LSTM still has disadvantages. The GA-LSTM does not consider the market sentiments. Therefore, we found another model, FinBert-LSTM. Finally, we think that only using the LSTM to predict the stock price's accuracy is not high, and FinBERT and LSTM together can increase prediction accuracy.

Acknowledgement

Shangyuan Liu, Weijian Chen, Lewei Tianand Junyi Zheng contributed equally to this work and should be considered co-first authors.

References

[1] Sha, Xinye. (2024) Time Series Stock Price Forecasting Based on Genetic Algorithm (Ga)-Long Short-Term Memory Network (LSTM) Optimization." arXiv.Org arxiv.org/abs/2405.03151
[2] YingZi. (2023) Stock market prediction based on the LSTM model in Python. (Python code) Csdn. https://download.csdn.net/download/ma_nong33/88351488
[3] Connor-LD. (2023). FinBERT Sentiment Analysis for Stock Prediction [Python code]. GitHub. https://github.com/Connor-LD/finBERT-Sentiment-Analysis-for-Stock-Prediction
[4] OkayJeff5. (2024). Stock price prediction using LSTM and Sentiment Analysis [Python code]. GitHub. https://github.com/OkayJeff5/Stock-price-prediction-using-LSTM-and-Sentiment-Analysis
[5] Nabipour, Mojtaba, et al. (2020) Deep Learning for Stock Market Prediction. arXiv.Org arxiv.org/abs/2004.01497