Forecasting Netflix Stock Prices by Time Series Analysis: ARIMA - LSTM Hybrid Model

Mingyan Liu

International college of Zhengzhou University, Zhengzhou University, Zhengzhou, Henan, 450000, China

lmy19839983910@stu.zzu.edu.cn

Abstract. With the intensifying competition in the streaming media market, the fluctuations in Netflix's stock price have attracted attention, and accurately predicting its stock price is of great significance to investors. Over the years, many studies have explored the use of time series analysis for stock price forecasting. This study uses the Autoregressive Integrated Moving Average and Long Short-Term Memory (ARIMA-LSTM) hybrid model to predict Netflix's stock price. It utilizes the stock price data from September 2002 to May 2024. The methodology involves using the ARIMA model to address linear trends, followed by the LSTM model to capture nonlinear characteristics. Findings indicate that this hybrid approach can effectively forecast stock prices. The result is more accurate than the ARIMA model. Future research can introduce more variables and pay attention to market changes to improve the model, providing references for investors and helping them optimize their investment portfolios and formulate risk-hedging strategies.

Keywords: time series analysis, stock price prediction, ARIMA-LSTM model.

1 Introduction

In the financial field, the stock market (SM) fluctuations significantly influence investors, decision-makers, and the entire economic system. Netflix, as a world-renowned streaming service provider, has pioneered a new era of digital media consumption since the late 1990s. However, over the past decade, with the emergence of an increasing number of streaming platforms, Netflix's stock price has witnessed a significant decline, attracting much attention. Based on this background, predicting its stock price has important practical significance. By improving predictive accuracy, investors can make better-informed decisions regarding their investments in Netflix, ultimately contributing to more effective risk management and capital allocation strategies [1]. Meanwhile, an in-depth exploration of the stock price trends of global enterprises like Netflix can provide references for research on the behavioral patterns of global SM.

As a statistical method for dealing with dynamic data, Time Series Analysis has unique advantages in stock price prediction. It can effectively explore the patterns of

W. Zang and C. Xia (eds.), *Proceedings of the 2025 3rd International Conference on Digital Economy and Management Science (CDEMS 2025)*, Advances in Economics, Business and Management Research 336,
https://doi.org/10.2991/978-94-6463-770-0_3

data changes over time and, through the analysis of historical data, identify characteristics such as trends, seasonality, and periodicity of stock prices (SP). Currently, many studies have applied time series analysis to the prediction of SP. Fakhfekh pointed out that in time series models for stock prediction, the Autoregressive Integrated Moving Average (ARIMA) models are the most widely used to predict future prices [2]. This model can capture the linear relationships in the data when dealing with stationary time series data. Brooks chose the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model to study the volatility of the SM and concluded that the model is used for one-step ahead forecasting, results show better prediction accuracy when compared with other models [3]. Besides, the research results of Samy et al. showed that Long Short-Term Memory (LSTM) networks have a lot of potential and show promise as an efficient solution to time series forecasting problems through the various sequences they have been applied to [4]. However, including the company's strategies, changes in user subscription behavior, fluctuations in the macroeconomic situation, and adjustments in policies, factors influencing Netflix's stock price are various, which makes its price changes rather complex. Therefore, it is not easy to make predictions using traditional single models.

In 1969, Bates first put forward the combined forecasting method which combines different models to analyze and predict data, so that the advantages of each model can be demonstrated while avoiding the disadvantages of each model at the same time [5]. Based on the analysis of existing research, this paper takes into account the advantages of the ARIMA model in capturing linear trends and the advantages of the LSTM model in dealing with nonlinear features and time dependence. By applying both ARIMA and LSTM models to forecast Netflix's stock price, this research aims to develop a more comprehensive prediction model that overcomes single-model limitations and expands the application of time series analysis in complex financial data prediction.

2 Method

2.1 Dataset

This study has collected the historical data of Netflix's SP from September 2002 to May 2024. The data is sourced from the Kaggle website and has relatively high validity and credibility. In addition, the closing price is an important reference date in the SM. It not only reflects the market sentiment of the day but also serves as a crucial factor for analyzing stock price trends. Consequently, in this project, the closing price is taken as the main data for modeling and prediction.

2.2 Model Introduction

ARIMA is a model for modeling linear dependencies in time series data [6]. This model adopts the differencing technique (first-order or multi-order) to transform non-stationary time series into stationary ones. The differencing term corresponds to the parameter d in the ARIMA model. Once the data becomes stationary, this model will build a combined autoregressive and moving average model based on this stationary sequence.

This model can be expressed as:

$$y_t' = c + \phi_1 y_{t-1}' + \cdots + \phi_p y_{t-p}' + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t \tag{1}$$

Among them, ${y_t}'$ is the stationary sequence after d times of differencing, c is the constant term, $\phi_1, \phi_2, \ldots, \phi_p$ is the parameter to be estimated in the autoregressive model, and $\theta_1, \theta_2, \ldots, \theta_q$ is the parameter to be estimated in the moving average model. $\epsilon_{t-q}, \ldots, \epsilon_{t-1}, \epsilon_t$ is the residual sequence of the model.

The LSTM model can capture the long-term dependence problem existing in sequences, that is, the current value of a sequence may be affected by an event that happened a long time ago [7]. The traditional ARIMA model cannot directly model such long-term dependence relationships in sequences, while the LSTM model can well store and remember such historical information through the Cell State and the Hidden State. The core of the LSTM model lies in the way of updating the Cell State $\tilde{C}_t$ and the Hidden State h_t.

The formula for the forget gate in the LSTM model is:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{2}$$

The update formula for the Cell State is:

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \tag{3}$$

The update formula for the hidden state is:

$$h_t = o_t \tanh(C_t) \tag{4}$$

2.3 Research Design

With the development of technology, traditional forecasting methods can no longer meet the needs of today's market. They can only rely on linear time series to predict SP and fluctuations while ignoring the nonlinear characteristics of stocks. This leads to the accuracy of the fitting model being affected and thus reduces the prediction precision [8]. To enhance the accuracy of forecasting Netflix's stock price trends, this project, based on Python, uses the ARIMA-LSTM hybrid model in time series to model and predict the target data. The process begins with data visualization, including a time series plot (Fig. 1) that reveals a clear trend in share prices.

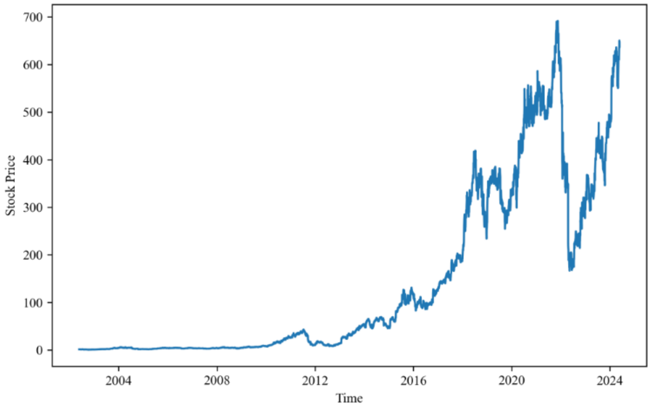

Fig. 1. Time series plot of Netflix's stock price (Photo credit: Original).

Before using the ARIMA model for modeling, it is necessary to perform a differencing operation on the time series. This paper uses the grid search method to determine the ARIMA model's p, d, and q parameters. Subsequently, the ARIMA model's residuals are computed and used as input for LSTM model training. The final prediction is obtained by combining the ARIMA and LSTM model predictions.

3 Result

It can be found in Table 1 that ARIMA (2, 1, 2) has the lowest AIC, BIC, and MAPE. Therefore, ARIMA (2, 1, 2) is used for modeling the stock price data in the subsequent part of this paper.

Table 1. Grid search results for ARIMA model parameters.

ARIMA Model	R2	Ljung-Box Statistic	AIC	BIC	P-value	MAPE
ARIMA (0,1,0)	0.9989	38.6613	35061.5850	35074.8060	0.1728	2.7708
ARIMA (1,1,0)	0.9989	37.2886	35061.6605	35081.4920	0.9906	2.7879
ARIMA (2,1,0)	0.9989	36.4092	35062.7672	35089.2092	0.9919	2.7767
ARIMA (0,1,1)	0.9989	37.3256	35061.7096	35081.5411	0.9861	2.7887
ARIMA (1,1,1)	0.9989	36.2070	35062.4346	35088.8766	0.9883	2.7813
ARIMA (2,1,1)	0.9989	37.3586	35065.6481	35098.7006	0.9875	2.7846
ARIMA (0,1,2)	0.9989	36.4055	35062.7505	35089.1925	0.9921	2.7779
ARIMA (1,1,2)	0.9989	37.3698	35064.3113	35097.3638	0.9968	2.7842
ARIMA (2,1,2)	0.9989	25.3615	35036.9040	35076.5670	0.6735	2.6969

Through Table 2, the parameter estimates, standard deviations, Z-statistics, p-values, and 95% confidence intervals of the ARIMA (2, 1, 2) model can be seen.

Among them, except for the intercept term, the parameter coefficients of the two AR terms and the two MA terms are all significantly different from 0. This indicates that the ARIMA (2, 1, 2) model can indeed capture the sequential dependence relationships in the stock price data of Netflix.

Table 2. Parameter information of the optimal ARIMA model

Parameter	coefficient	std err	z	P>|z|	[0.025	0.975]
Intercept	0.1475	0.1020	1.4450	0.1480	-0.0530	0.3480
ar.L1	0.6482	0.0200	31.8460	0.0000	0.6080	0.6880
ar.L2	-0.8776	0.0170	-52.9650	0.0000	-0.9100	-0.8450
ma.L1	-0.6780	0.0180	-37.3250	0.0000	-0.7140	-0.6420
ma.L2	0.9087	0.0150	61.0660	0.0000	0.8790	0.9380
sigma2	35.6955	0.1330	268.8330	0.0000	35.4350	35.9560

Table 3 presents the Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) indicators of the ARIMA model and the ARIMA - LSTM hybrid model on the fitting data and the test data. During the fitting stage, the ARIMA-LSTM model shows slightly higher MAE, RMSE, and MAPE compared to the ARIMA model. Conversely, in the prediction stage, the ARIMA-LSTM model exhibits lower values across these indicators.

Table 3. Comparison of model ARIMA and model ARIMA-LSTM predictions

Model	Period	MAE	RMSE	MAPE
ARIMA	Fitted	2.5431	5.9238	0.0256
ARIMA	Prediction	110.2148	122.7704	0.1862
ARIMA_LSTM	Fitted	2.8680	6.0248	0.0632
ARIMA_LSTM	Prediction	109.4101	122.1434	0.1847

Finally, Fig. 2 displays the visual results of the ARIMA-LSTM hybrid model after the training set and the test set are divided, as well as the prediction results of the stock price changes in the next year. It can be seen that when the ARIMA-LSTM hybrid model conducts multi-step predictions, the confidence interval will gradually increase as the number of prediction steps increases, which is a typical characteristic of autoregressive models.

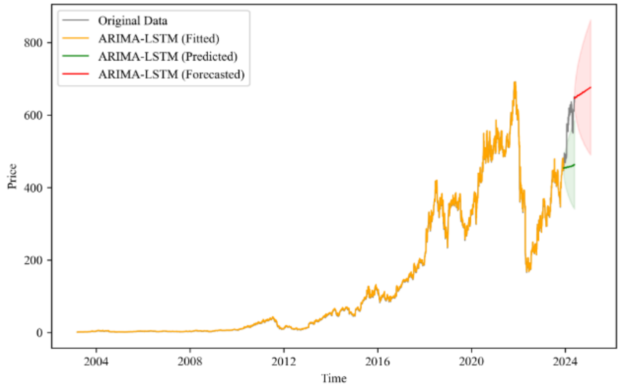

Fig. 2. SP of Netflix predicted based on the ARIMA-LSTM hybrid model (Photo credit: Original).

4 Discussion

4.1 Result Explanation

The findings of this study demonstrate that the ARIMA-LSTM model effectively captures Netflix stock price trends and characteristics, enabling future price predictions. Through experiments, it is found that the stock price shows a slight upward trend in the coming period. The model's strong performance on both training and validation sets suggests its predictive capabilities, offering valuable insights for investors to better grasp the possible changing trends of SP in investment decisions.

Yuan et al. believed that machine learning models can take nonlinear features as inputs and can be used to predict non-stationary financial time series, such as stocks, futures, digital currencies, etc [9]. This method has higher accuracy and smaller errors compared with traditional time series prediction methods. The specific parameters obtained in this study just verify this view. It is found that the combined use of ARIMA and LSTM models can indeed improve the prediction accuracy compared with using only the ARIMA model.

4.2 Limitation Analysis

Although certain achievements have been made in using the ARIMA-LSTM model to predict data, there are still some limitations. Firstly, it cannot be ignored that the SM is renowned for its unpredictability, nonlinearity, and active essence [10]. The model used in the study cannot predict every fluctuation of SP in the future and can only make simple predictions on the overall trend.

In addition, during the data fitting stage, the ARIMA-LSTM model exhibits a slightly higher error than the ARIMA model, which is unexpected in the study. This could be because the ARIMA-LSTM hybrid model is relatively complex and prone to overfitting. Especially for the LSTM part, if the amount of training data is limited, it may overfit the noise and fluctuations in the training data, resulting in relatively large errors.

4.3 Suggestions

Based on the above content, this project puts forward a series of suggestions to improve future research and applications. Firstly, in terms of model construction, more external variables can be introduced, such as macroeconomic indicators, user sentiment analysis data, etc., so that the model can reflect the actual situation of the SM more comprehensively. In addition, continuously paying attention to the new features and changes in the SM and adjusting the model algorithm on time can improve prediction accuracy and consistency.

For investors, the modeling results can only serve as references and should not be the sole basis for investment decisions. When formulating investment strategies, investors should comprehensively consider the fundamental analysis of companies, including financial conditions, business models, market competitiveness, etc., to evaluate investment risks and potential returns. On the other hand, investors should closely follow market dynamics, continuously improve their risk awareness, and make rational investments.

5 Conclusion

This study initially uses the ARIMA model to analyze Netflix's SP and extract residuals. Subsequently, the LSTM model is trained on these residuals, utilizing historical price data to capture nonlinear fluctuations unexplained by the ARIMA model. After constructing the model, the data of the most recent 116 days of Netflix's SP are reserved to test prediction results. Meanwhile, the SP data for the whole next year (252 days) are predicted by the hybrid model. While the ARIMA-LSTM hybrid model demonstrates slight improvements in prediction accuracy compared to the standalone ARIMA model, the enhancements are limited for stock price time series. In the future, besides providing investors with predictive references for stock price trends, the research findings can also be applied to the field of financial risk management. Financial institutions can utilize the model's prediction to assess financial portfolios more accurately and optimize risk hedging strategies.

References

1. Hanif, M., Thomas: Best Forecasting Netflix SP: A Case Study Using Regression and Machine Learning Models. ResearchGate, October 2024.
2. Fakhfekh, M., Jeribi, A.: Volatility dynamics of cryptocurrencies' returns: Evidence from asymmetric and long memory GARCH models. Research in International Business and Finance 51, 101075 (2020).
3. Brooks, C.: Predicting stock index volatility: Can market volume help? Journal of Forecasting 17(1), 59–80 (1998).

4. Bengio, S., Vinyals, O., Jaitly, N., Shazeer, N.: Scheduled Sampling for Sequence Prediction with Recurrent Neural Networks. Advances in Neural Information Processing Systems 28 (NIPS 2015), Mountain View, CA, USA (2015).
5. Bates, J.M., Granger, C.W.: The combination of forecasts. Journal of the Operational Research Society 20(4), 451–468 (1969).
6. Autoregressive Integrated Moving Average - an overview. ScienceDirect Topics, https://www.sciencedirect.com/topics/mathematics/autoregressive-integrated-moving-average (2024-12-16).
7. Hochreiter, S., Schmidhuber, J.: Long Short-Term Memory. Neural Comput. 9(8), 1735–1780 (1997).
8. Sunny, M., Maswood, M., Alharbi, A.G.: Deep Learning-Based Stock Price Prediction Using LSTM and Bi-Directional LSTM Model. In: Proceedings of the 2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES), 2020.
9. Yuan, W., Chin, K.S., Hua, M., et al.: Shape classification of wear particles by image boundary analysis using machine learning algorithms. Mechanical Systems and Signal Processing 14(5), 72–73 (2016).
10. Vijh, M., Chandola, D., Tikkiwal, V.A., Kumar, A.: Stock closing price prediction using machine learning techniques. Procedia Computer Science 167, 599–606 (2020).