ISSN 2616-5902 Vol. 5, Issue 8: 38-43, DOI: 10.25236/AJBM.2023.050807

Comparing the ARIMA and LSTM Models on the Stock Price of FinTech Companies

Ying Zhang[1,a,*], Zixin Tong[1,b]

[1]*Wenzhou-Kean University, No. 88, University Road, Li'ao Street, Ouhai District, Wenzhou, Zhejiang, China*
[a]*zhying@kean.edu*, [b]*tongzix@kean.edu*
[*]*Corresponding author*

Abstract: *FinTech companies have emerged as a new force in the financial industry in recent years. Their innovative business models, high growth performance, and broad market prospects have attracted the attention and pursuit of numerous investors. Therefore, predicting the future trend of FinTech company stocks is significant for investors. This paper selects PayPal as the prediction target, collects the closing prices from 2018 to 2023, and uses the Auto-regressive Integrated Moving Average Model (ARIMA) and Long Short-Term Memory (LSTM) to compare the accuracy of the prediction results and select the optimal prediction model. Empirical results show that both the ARIMA and the LSTM models have specific effects in predicting the stock prices of the FinTech industry, with the LSTM model having higher prediction accuracy than the ARIMA model. However, it should be noted that stock price prediction is not entirely accurate but based on historical data and market trends. Therefore, when making investment decisions, it is necessary to consider various factors comprehensively and make scientific investment decisions to achieve better investment returns.*

Keywords: *FinTech; PayPal; ARIMA model; LSTM model; Stock trend*

1. Introduction

FinTech companies are an important force in the global financial market and play an essential role in the innovation and development of the financial industry. It involves digital payments, blockchain technology, lending and financing, insurance technology, artificial intelligence and extensive data analysis, e-commerce, and retail payments [1]. Some well-known FinTech companies include PayPal Holdings, Inc. (PYPL), Square, Inc. (SQ), Shopify Inc. (SHOP), Ant Group Co., Ltd., etc. These companies have achieved great success in different fields or markets, with a certain influence and driving the development and transformation of global financial technology.

With the continuous development and popularization of the FinTech industry, more people are paying attention to the stock performance of FinTech companies. Stocks of many FinTech companies have outperformed against the backdrop of digital transformation accelerated by the global COVID-19 epidemic. Taking the United States as an example, during the 2020 pandemic, digital payments surged, with the stock price of Square (SQ) rising by more than 250% and PayPal (PYPL) rising by over 110% [2]. The market recognizes the performance and prospects of these companies, and investors are confident in their prospects. However, due to the rapid development of the FinTech industry, investing in FinTech companies also carries certain risks. Therefore, before investing in FinTech stocks, investors need to conduct a comprehensive analysis and research on the company's business model, profitability, competitive advantage, and other aspects, and do an excellent job in risk control.

To predict the future development trend of the stock market and the degree of stock price fluctuations, time sequence models, regression models, and integrated model artificial neural networks are widely used by researchers. Among them, time sequence models include ARIMA, VAR, and LSTM, which are based on historical data to establish models to predict future stock price changes. This article uses ARIMA and LSTM models to analyze the stock growth trend of the FinTech industry, and wants to understand the stock price change trend and the investment value of PayPal (one of the world's largest digital payment platforms) before and after COVID-19. It is planned to compare the applicability of the two models in the study of stock market fluctuations, select a better forecasting model, and help investors control risks.

The remainder of this paper is organized as follows. Section 2 summarizes existing studies on applying ARIMA and LSTM in the stock market. In Section 3, we provide an overview of experimental methods. The detailed analysis and demonstration results of the experimental process are discussed in Section 4. The paper is concluded with Section 5.

2. Literature Review

In recent years, the FinTech industry has developed rapidly, especially during the epidemic, where its importance has become more prominent. Some studies show that the epidemic has accelerated the Digital transformation and development of the FinTech industry, while also bringing challenges and opportunities to the industry [3]. Many researchers have started to apply time-sequence models to analyze the stock data of FinTech companies to understand their trends and predict future developments.

ARIMA and LSTM are standard methods among them. The LSTM model has good performance in predicting stock prices in the financial market and has good tracking ability [4]. Meanwhile, the LSTM model can effectively predict the success rate of FinTech companies [5]. Traditionally, the ARIMA model is one of the most widely used linear models in time series forecasting, but it cannot effectively depict nonlinear patterns [6]. Some literature compares ARIMA and LSTM models. According to the research conducted by Xiao and Feng in 2022, both the ARIMA model and the LSTM model can predict stock prices, and the predicted results are basically consistent with the actual results. LSTM performs well in predicting stock prices (especially expressing changes in stock prices), while ARIMA models are more convenient to apply [7]. In summary, LSTM models usually perform better than ARIMA in stock price forecasting, especially for nonlinear and non-stationary data. Moreover, the predictive performance of LSTM models can be further improved when choosing proper hyperparameters [8]. However, since each market and stock have different characteristics, the application of these models for forecasting requires evaluation and adjustment to specific circumstances.

Many scholars have applied ARIMA and LSTM to analyze FinTech companies. ARIMA and LSTM models have great potential in financial technology stock analysis, and combining multiple data sources and technologies can improve prediction accuracy [9]. Some scholars combine ARIMA and LSTM to predict the future market value of FinTech companies in the United States, and incorporate data preprocessing techniques such as differentiation, normalization, and normalization to improve the accuracy of predictions. However, the prediction results have limitations. Porta and Facchinetti developed a hybrid model in 2019 that combines ARIMA and LSTM models. The performance of hybrid models is superior to independent ARIMA and LSTM models, indicating the potential of these combined models in enhancing FinTech stock analysis. However, complex calculations, feature engineering, and model selection pose significant challenges [10]. These findings are of great significance for understanding potential trends and promoting informed predictions of the future trajectory of the emerging FinTech industry.

Based on existing research, this article applies ARIMA and LSTM models to the stock price prediction of the FinTech industry, with PayPal company as the representative, and compares and selects a better prediction model.

3. Methodology

3.1. Auto-regressive Integrated Moving Average Model (ARIMA)

A generalized version of the Auto-regressive Moving Average (ARMA), the ARIMA model integrates the Auto-regressive (AR) and Moving Average (MA) processes to create a composite model of the time series [11]. The general form of the ARIMA model is expressed as ARIMA (p, d, q). Where p denotes the auto-regressive parts of the data set, d refers to integrated parts of the data set, q represents moving average parts, and p, d, q is all non-negative integers.

ARIMA (p, d, q) captures the model’s key elements. A simple form of an AR model of order p, i.e., AR(p), can be expressed as a linear process by:

$$x_t = c + \sum_{i=1}^{p} \Phi_i x_{t-i} + \epsilon_t$$

x_t is a stationary variable, and c is a constant, where the term in Φ_i is auto-correlation coefficients at lags 1, 2, p, and ϵ_t. The residuals have a mean of zero and a variance σ_{ϵ}^2, and are a series of Gaussian white noise. An MA model of order q, often known as MA(q), can be written in the form:

$$x_t = \mu + \sum_{i=0}^{q} \Theta_i \epsilon_{t-i}$$

Here μ is the expectation of x_t, which is usually assumed to be 0. The θ_i term is the weights applied to a stochastic term's current and prior values in the time series, $\theta_0 = 1$. With zero mean and variance σ_2, we assume that t is a Gaussian white noise sequence. These two models can be combined by adding them up to create an ARIMA model of order (p, q):

$$x_t = c + \sum_{i=1}^{p} \Phi_i x_{t-i} + \epsilon_t + \sum_{i=0}^{q} \Theta_i \epsilon_{t-i}$$

In this equation, $\Phi_i \neq 0$, $\theta_i \neq 0$, and $\sigma_{\epsilon}^2 > 0$.

In order to determine the optimal ARIMA model in multiple experiments, this study adopts the following four criteria for stock indices. Firstly, the relatively small BIC (Bayesian or Schwartz Information Criterion). Secondly, the relatively small standard error of the regression (S.E. of regression). Thirdly, the adjusted R^2. Fourthly, relatively high Q-statistics and correlation graphs show no obvious pattern left in the auto-correlation function (ACF) and partial auto-correlation function (PACF) of the residual, which means that the residual of the selected model is white noise.

3.2. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a kind of Recurrent Neural Network (RNN) capable of remembering the values from earlier stages for future use [12]. The LSTM algorithm can capture long-distance dependencies, and the gradient does not disappear. A classic LSTM cell and an LSTM cell chain are shown in Figure 1.

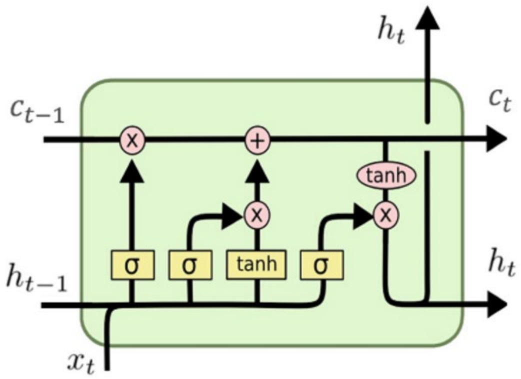

Figure 1: Structure of the LSTM cell

Figure 1 shows that the state of each LSTM cell can be divided into two vectors, short-term state and long-term state, controlled by three gates: forgetting gate, input gate, and output gate. The forgetting gate controls which part of the previous time step state c (t-1) should be discarded in this time step, and the input gate determines σ Which part of the should be added to the long-term state, and the output gate selects which part of the long-term state should be output to ht. At each time step, some memory is kept, and some memory is added, so LSTM is of great help in solving gradient vanishing problems and even allows the neural network to learn long dependencies. Our case aims to analyze whether the LSTM model can accurately predict the stock price of PayPal Corporation. This is a time-series-dependent problem, and information about previous company stock price changes will help predict the outcome of the next time step.

4. Experimental Analysis

4.1. Data Selection

In this study, there is a total of 1324 data for PayPal company, collected from Yahoo Finance,

ISSN 2616-5902 Vol. 5, Issue 8: 38-43, DOI: 10.25236/AJBM.2023.050807

beginning from January 1, 2018 to April 7, 2023. The stock data contains the opening price, low price, high price, closing price, adjusted closing price, and volume. This paper uses the closing price to represent the activities in one day. Among the 1324 data, there are 1060 (80%) for training and 264 (20%) for testing.

There is a peak starting from mid-2020 to mid-2021 of the stock price of PayPal. The COVID-19 pandemic seems to have contributed to digitalization [13], and since PayPal is a digital payments company, the frequent use of online payments will benefit its customer confidence, and further foster its stock price. No sooner after COVID-19 spreads all over the world, the stock price goes up and after 2022, it begins to return to normal, as visualized in Figure 2.

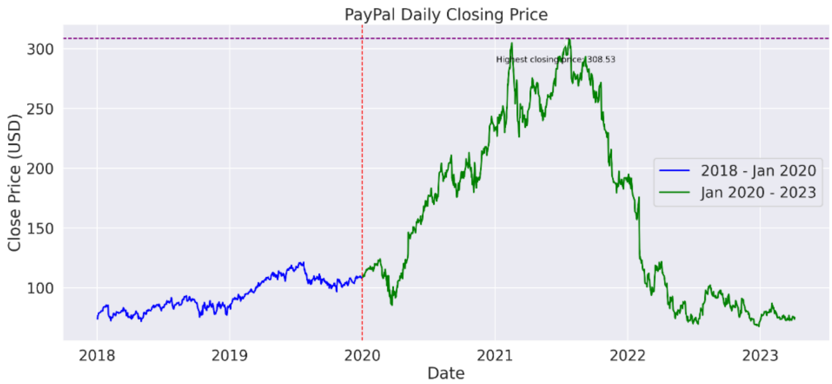

Figure 2: The Stock price of PayPal from 2018 to present

4.2. ARIMA Model for Stock Price of PayPal Holdings, Inc.

4.2.1 Stabilization Treatment and Model Fitting

To begin with, the Stationarity test is required for time series analysis. Applying ARIMA on non-stationary data can lead to unreliable or spurious results. According to Table 1, the ADF (Augmented Dickey-Fuller) test has a result of -1.05 when the p-value is 0.74, which is greater than 0.05, meaning the data is not static. In order to stabilize the data, the difference is processed. After the process, the p-value is smaller than 0.05. Then, ARIMA can be applied later.

Table 1: ADF test before and after the difference

<table>
<tr><th>Test</th><th>ADF test statistic</th><th>p-value</th><th colspan="3">Critical values</th></tr>
<tr><td>Before</td><td>-1.05</td><td>0.74</td><td rowspan="2">1%:
-3.44</td><td rowspan="2">5%:
-2.86</td><td rowspan="2">10%:
-2.57</td></tr>
<tr><td>After</td><td>-36.27</td><td>0.00</td></tr>
</table>

Then, from Figure 3, it can be found that the autocorrelation coefficient and the partial correlation coefficient basically lie on the same level, meaning they have the same confidence interval. Therefore, this paper applies ARIMA (1,1,1) for the series.

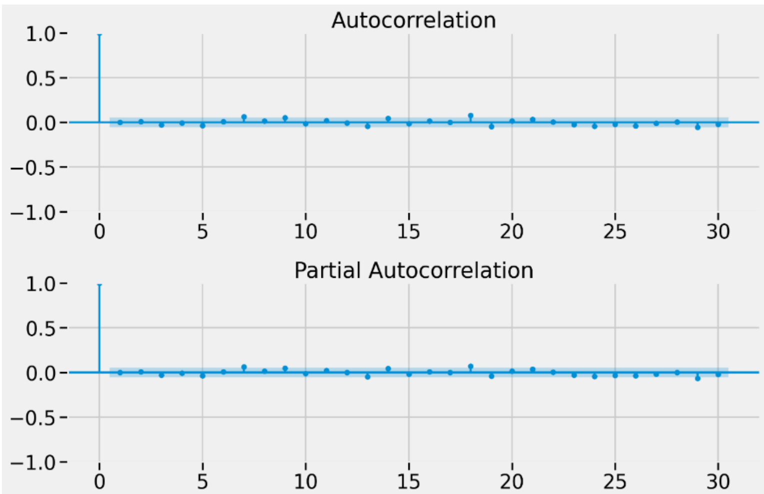

Figure 3: Autocorrelation graph (ACF) and Partial Autocorrelation graph (PACF)

ISSN 2616-5902 Vol. 5, Issue 8: 38-43, DOI: 10.25236/AJBM.2023.050807

4.2.2 Stabilization Treatment and Model Fitting

From Figure 4, the residuals plot of ARIMA (1,1,1), the residual sequence is random white noise sequence, proving the efficiency of the model. The analysis of this ARIMA (1,1,1) is reasonable and effective. From Figure 5, based on the testing data, it is obvious that the predicted future stock value is good fitting the actual future stock value.

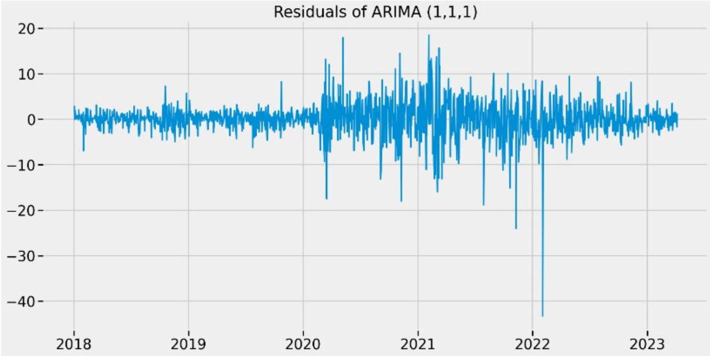

Figure 4: Residuals plot of ARIMA (1,1,1) Model

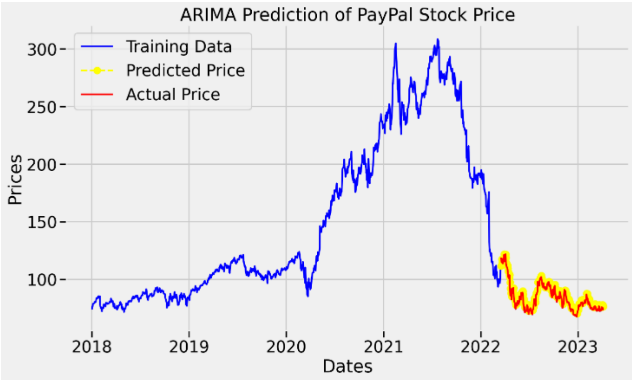

Figure 5: ARIMA prediction of PayPal stock price

4.3. LSTM Model for Stock Price of PayPal Holdings, Inc.

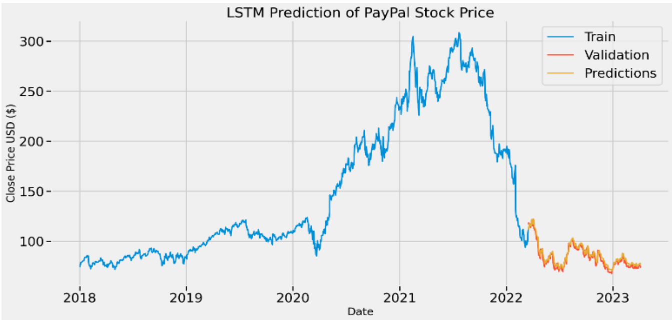

Figure 6: ARIMA prediction of PayPal stock price

This paper builds an LSTM (Long Short-Term Memory) model, which is capable of learning long-term dependencies in sequence data. The first LSTM layer has 50 units, then add the second layer with 50 units, and then two Dense (fully connected) layers which each has 25 units, and whose the second Dense layer has a single unit, corresponding to the predicted value.

From Figure 6, we can visualize the prediction with respect to the time of years. The price mainly fluctuates from around 100 to 300, and there is a short peak after COVID-19, starting in 2020.

4.4. Comparison of ARIMA and LSTM

RMSE (root mean square error) can be used to analyze the accuracy of the models. It computes the

ISSN 2616-5902 Vol. 5, Issue 8: 38-43, DOI: 10.25236/AJBM.2023.050807

differences or residuals between the tested and predicted values. The indicator equation is as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2}$$

According to Table 2, the RMSE for ARIMA and LSTM are 3.37 and 2.02 respectively. This indicates that LSTM has fewer errors and has better performance than ARIMA for predicting the future stock price of PayPal company.

Table 2: Error for ARIMA and LSTM

Model	MSE	RMSE
ARIMA	11.37	3.37
LSTM	4.07	2.02

5. Conclusions

This paper uses two predicting machine learning models, ARIMA and LSTM, to predict PayPal closing stock prices from 2018 to 2023. After preprocessing the data, selecting appropriate model parameters, and training both models, the predictions were evaluated using the RMSE metric. During this period, LSTM has better performance in PayPal stock price forecasting, achieving a lower RMSE value. With the unexpected influence of COVID-19 in 2020, the LSTM can better deal with the emergency. On the other hand, it is crucial to note that stock price prediction is inherently difficult due to market volatility, macroeconomic events, and unexpected company news. Thus, better models should be designed to face these factors to improve prediction accuracy in the future.

References

[1] Trificana J. "What is FinTech? 6 Main Types of FinTech and How They Work." Plaid, 2022, https://plaid.com/resources/FinTech/what-is-FinTech/.

[2] Gerrans Paul, et al. "FinTech and Responsibility: Buy-now-pay-later Arrangements." Australian Journal of Management, vol. 47, no. 3, SAGE Publications, July 2021, pp. 474–502.

[3] Akhtaruzzaman M., Boubaker S., & Sensoy A. (2021, January). Financial contagion during COVID–19 crisis. Finance Research Letters, 38, 101604.

[4] Moghar A., & Hamiche,M. (2020). Stock Market Prediction Using LSTM Recurrent Neural Network. Procedia Computer Science, 170, 1168–1173.

[5] Vochozka M., Vrbka J., & Suler P. (2020, September 12). Bankruptcy or Success? The Effective Prediction of a Company's Financial Development Using LSTM. Sustainability, 12(18), 7529.

[6] Pai P. F., & Lin C. S. (2005, December). A hybrid ARIMA and support vector machines model in stock price forecasting. Omega, 33(6), 497–505.

[7] Wang J., Yang Q., & Jin Z. (2022). Research on Stock Forecasting Based on LSTM-ARIMA Model. Advances in Social Sciences, 11(07), 2843–2856. https://doi.org/10.12677/ass.2022.117390

[8] Aszemi N. M., & Dominic P. (2019). Hyperparameter Optimization in Convolutional Neural Network using Genetic Algorithms. International Journal of Advanced Computer Science and Applications, 10(6).

[9] Ho M. K., Darman H., & Musa S. (2021, July 1). Stock Price Prediction Using ARIMA, Neural Network and LSTM Models. Journal of Physics: Conference Series, 1988(1), 012041. https: // doi. org/ 10. 1088/1742-6596/1988/1/012041

[10] Porta M. & Facchinetti T. (2019). Applied Data Science Approaches in FinTech: Innovative Models for Bitcoin Price Dynamics.

[11] Hamilton James Douglas. Time Series Analysis. 1994. Bowker, https: // doi. org/ 10. 1604/ 9780691042893.

[12] Xie Jiaheng, et al. "Understanding Medication Nonadherence from Social Media: A Sentiment-Enriched Deep Learning Approach." SSRN Electronic Journal, Elsevier BV, 2017.

[13] Nanda A., Xu, Y., & Zhang, F. (2021). How would the COVID-19 pandemic reshape retail real estate and high streets through acceleration of E-commerce and digitalization? Journal of Urban Management, 10(2), 110-124.