Predicting Stock Prices Using ARIMA and LSTM

University of Southern California
Qianran Ma (7798937244)

Abstract

In this report, I collected five-year stock price data of NVIDIA and Microsoft to make a prediction of future stock prices. I built a statistic model, ARIMA model, and a deep learning model, LSTM model to solve the problem. After tuning the model, I find the best (p, q) for NVDA ARIMA model is (5, 1), the best (p, q) for MSFT ARIMA model is (3, 2), and the RMSE of best NVDA LSTM is 15.43, the RMSE of best MSFT LSTM is 9.40. Also, I find that ARIMA can only be used to predict short-term price, while LSTM is able to predict long-term price.

1 Introduction

The stocks of two technology companies, which are NVIDIA (Stock: NVDA) and Microsoft (Stock: MSFT), are selected for this report. NVIDIA is a global leading company which designs hardware and software for artificial intelligence research and application. Microsoft Corporation is also a world famous technology corporation producing computer software, consumer electronics, personal computers, and related services.

This report covers the data collection of two stocks, two solutions to make stock price predictions, including using ARIMA model and LSTM model, then compares the two models, and explains its financial value.

2 Data

The closing prices data from 2017-12-1 to 2022-11-30 of NVIDIA's stock and Microsoft's stock are collected, with a total number of 2516. There are no missing data and outliers in the dataset. The prices of two stocks both showed an increasing tendency from 2013 to 2022, and a decreasing tendency from 2022 to 2023. The price of Microsoft is higher than that of NVIDIA in average, but their maximum price is similar.

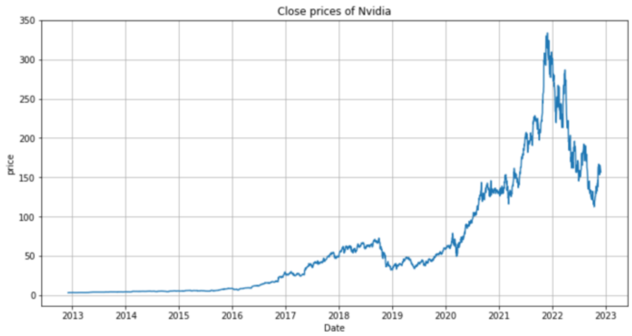

Fig. 1. Closing prices of NVDA data from 2017-12-1 to 2022-11-30

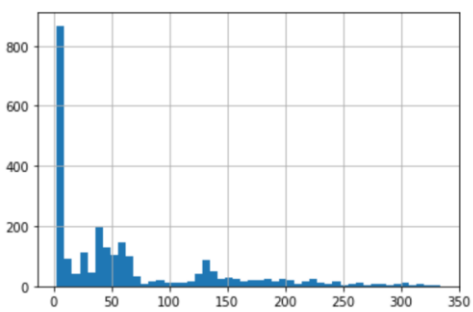

Fig. 2. The price distribution of NVDA

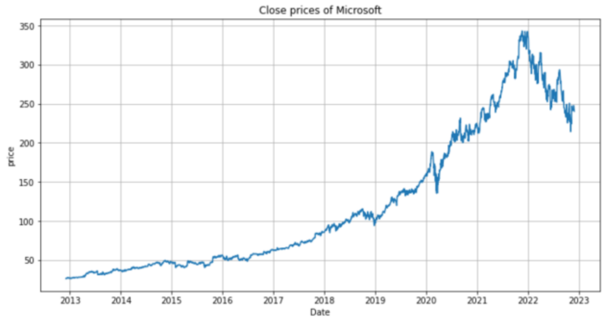

Fig. 3. Closing prices of MSFT data from 2017-12-1 to 2022-11-30

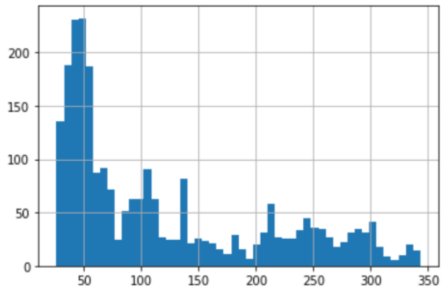

Fig. 4. The price distribution of MSFT

	NVDA	MSFT		NVDA	MSFT
count	2516	2516	**25%**	5.52	46.75
mean	63.29	119.98	**50%**	39.34	84.15
std	73.25	89.53	**75%**	88.11	186.86
min	2.94	26.37	**max**	333.76	343.11

Table 1. Data description of NVDA and MSFT

3 ARIMA Solution

ARIMA is a class of time series prediction models, and the name is an abbreviation for AutoRegressive Integrated Moving Average. The backbone of ARIMA is a mathematical model that represents the time series values using its past values. This model is based on two main features: past values, which is a good predictor of the future, and past errors, which shows how well it has performed in the past. Therefore, the model has two hyperparameters: p (the last p time series values) and q (the most recent q errors).

3.1 Data Preparation

ARIMA model requires the data to be stationary, which means the time series' basic characteristics remain unchanged. Only when a time series is stationary, can we predict the future features based on previous or current features. Therefore, some data preprocessing and stationary test are need before model building.

3.1.1 Stationary Test

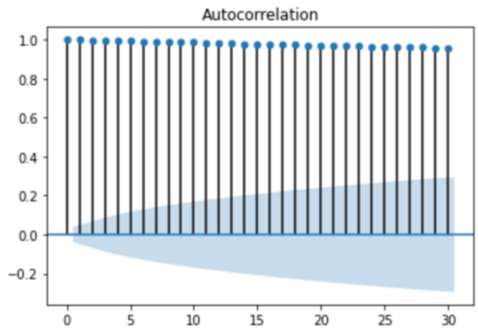

Fig. 5. ACF plot of NVDA price data

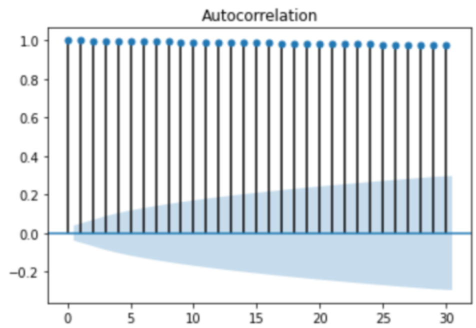

Fig. 6. ACF plot of MSFT price data

Fig 5 and Fig 6 show that the raw prices data of NVDA and MSFT did not meet the requirement of stationary sequence. Therefore, I converted the data into first difference.

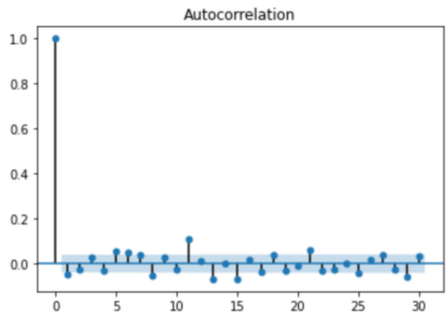

Fig. 7. ACF plot of NVDA first difference

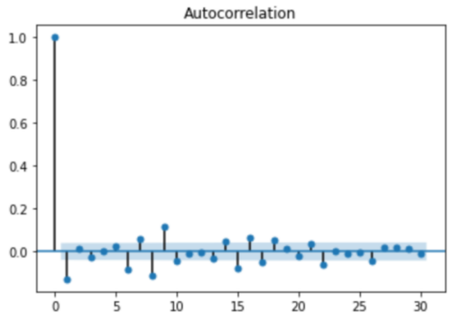

Fig. 8. ACF plot of MSFT first difference

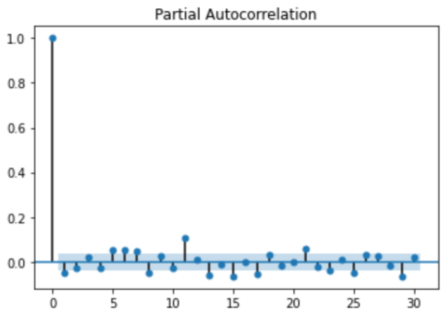

Fig. 9. PACF plot of NVDA first difference

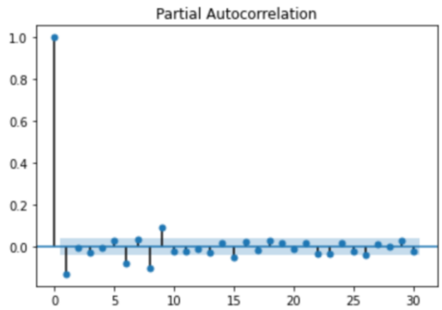

Fig. 10. PACF plot of MSFT first difference

```
   Augmented Dickey-Fuller Results
=====================================
Test Statistic                -10.617
P-value                         0.000
Lags                               20
-------------------------------------

Trend: Constant
Critical Values: -3.43 (1%), -2.86 (5%), -2.57 (10%)
Null Hypothesis: The process contains a unit root.
Alternative Hypothesis: The process is weakly stationary.
```

Fig.11. ADF result of NVDA first difference

```
   Augmented Dickey-Fuller Results
=====================================
Test Statistic                -17.132
P-value                         0.000
Lags                                8
-------------------------------------

Trend: Constant
Critical Values: -3.43 (1%), -2.86 (5%), -2.57 (10%)
Null Hypothesis: The process contains a unit root.
Alternative Hypothesis: The process is weakly stationary.
```

Fig.12. ADF result of MSFT first difference

From the ACF plot and PACF plot, we can see that first differences of NVDA and MSFT price data are stationary. The ADF results of them are much smaller than 0.05, which also proves they are stationary.

3.1.2 White Noise Test

I conducted a LB test to judge whether the first differences of NVDA and MSFT price are white noise since the ARIMA model cannot be built on white noise sequence. The LB results of them are 9.86e-11 and 1.05e-25, which are both much smaller than 0.05. Therefore, I can reject the null hypothesis and conclude that they are not white noise sequence.

3.1.3 Data Splitting

The final data with the size of 2515 is split into two datasets for the purpose of training and testing. I split 90% of the data as train set with the size of 2263, and 10% as the test set with the size of 252.

```
   Augmented Dickey-Fuller Results
=====================================
Test Statistic                -10.820
P-value                         0.000
Lags                               10
-------------------------------------

Trend: Constant
Critical Values: -3.43 (1%), -2.86 (5%), -2.57 (10%)
Null Hypothesis: The process contains a unit root.
Alternative Hypothesis: The process is weakly stationary.
```

Fig.13. ADF result of NVDA train set

```
   Augmented Dickey-Fuller Results
=====================================
Test Statistic                -14.830
P-value                         0.000
Lags                                8
-------------------------------------

Trend: Constant
Critical Values: -3.43 (1%), -2.86 (5%), -2.57 (10%)
Null Hypothesis: The process contains a unit root.
Alternative Hypothesis: The process is weakly stationary.
```

Fig.14. ADF result of MSFT train set

The training data also met the requirement of stationary sequence.

3.2 Model Building

I tried different p ranging from 0 to 5 and different q ranging from 0 to 2 to train ARIMA model. Akaike information criterion (AIC) and the Bayesian information criterion (BIC) provide measures of model performance that account for model complexity. They combine a term reflecting how well the model fits the data with a term that penalizes the model in proportion to its number of parameters. Therefore, I used these two criterion to evaluate the model.

(p, q)	AIC	BIC	Average of AIC and BIC
0,0	9880.30	9891.75	9886.03
0,1	9869.40	9886.57	9877.98
0,2	9864.07	9886.97	9875.52
1,0	9868.08	9885.25	9876.66
1,1	9867.93	9890.83	9879.38
2,0	9865.31	9888.21	9876.76
2,1	9842.89	9871.51	9857.20
2,2	9846.00	9880.34	9863.17
3,0	9865.19	9893.82	9879.50
3,1	9860.14	9894.49	9877.32
3,2	9842.16	9882.23	9862.19
4,0	9861.63	9895.97	9878.80
4,1	9858.62	9898.69	9878.66
4,2	9856.85	9902.64	9879.74
5,0	9847.43	9887.50	9867.47
5,1	**9828.82**	**9874.61**	**9851.72**
5,2	9830.75	9882.27	9856.51

Table 2. Tune results of ARIMA on NVDA data

(p, q)	AIC	BIC	Average of AIC and BIC
0,0	9873.82	9885.27	9879.55
0,1	9783.02	9800.20	9791.61
0,2	9775.48	9798.38	9786.93
1,0	9774.58	9791.75	9783.16
1,1	9775.31	9798.21	9786.76
1,2	9785.19	9813.81	9799.50
2,0	9775.17	9798.07	9786.62
2,1	9771.23	9799.85	9785.54
2,2	9733.06	9767.41	9750.24
3,0	9777.03	9805.65	9791.34
3,1	9764.38	9798.72	9781.55
3,2	**9716.36**	**9756.43**	**9736.39**
4,0	9777.87	9812.21	9795.04
4,1	9761.58	9801.65	9781.62
4,2	9718.31	9764.10	9741.20
5,0	9779.83	9819.90	9799.87
5,1	9759.31	9805.11	9782.21
5,2	9717.82	9769.34	9743.58

Table 3. Tune results of ARIMA on MSFT data

Take consideration of both AIC and BIC criterion, I calculated the average of AIC and BIC, and selected the best model with the lowest average of AIC and BIC. The parameters of best ARIMA model for NVDA is p = 5, q = 1; the parameters of best ARIMA model for MSFT is p = 3, q = 2.

3.3 Model Diagnosis

The information of the best ARIMA model for NVDA data and for MSFT data are shown as fig.15 and fig.16. The p-value of AR1, AR2, AR4, AR5, and MA1 in ARIMA(5, 1, 1) are all smaller than 0.05, and their confidence intervals do not include 0, so that these parameters are significant. The p-value of AR1, AR2, AR3, MA1, and MA2 are also smaller than 0.05, and their confidence intervals do not include 0, so that these parameters are significant.

ARMA Model Results

Dep. Variable:	Close	No. Observations:	2263
Model:	ARMA(5, 1)	Log Likelihood	-4906.409
Method:	css-mle	S.D. of innovations	2.115
Date:	Sun, 11 Dec 2022	AIC	9828.818
Time:	23:23:04	BIC	9874.613
Sample:	0	HQIC	9845.528

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.1510	0.063	2.406	0.016	0.028	0.274
ar.L1.Close	0.7060	0.068	10.393	0.000	0.573	0.839
ar.L2.Close	0.1065	0.027	3.980	0.000	0.054	0.159
ar.L3.Close	-0.0119	0.027	-0.445	0.656	-0.064	0.040
ar.L4.Close	-0.0699	0.027	-2.616	0.009	-0.122	-0.018
ar.L5.Close	0.1208	0.023	5.347	0.000	0.077	0.165
ma.L1.Close	-0.7903	0.066	-12.029	0.000	-0.919	-0.662

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	-1.2550	-1.0760j	1.6531	-0.3872
AR.2	-1.2550	+1.0760j	1.6531	0.3872
AR.3	1.1127	-0.0000j	1.1127	-0.0000
AR.4	0.9878	-1.3215j	1.6499	-0.1478
AR.5	0.9878	+1.3215j	1.6499	0.1478
MA.1	1.2653	+0.0000j	1.2653	0.0000

Fig.15. The information of ARIMA(5,1,1) for NVDA

ARMA Model Results

Dep. Variable:	Close	No. Observations:	2263
Model:	ARMA(3, 2)	Log Likelihood	-4851.179
Method:	css-mle	S.D. of innovations	2.064
Date:	Mon, 12 Dec 2022	AIC	9716.358
Time:	00:21:53	BIC	9756.430
Sample:	0	HQIC	9730.980

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.1367	0.037	3.685	0.000	0.064	0.209
ar.L1.Close	-1.9651	0.046	-42.296	0.000	-2.056	-1.874
ar.L2.Close	-1.2052	0.067	-18.074	0.000	-1.336	-1.075
ar.L3.Close	-0.1358	0.029	-4.687	0.000	-0.193	-0.079
ma.L1.Close	1.7874	0.040	44.651	0.000	1.709	1.866
ma.L2.Close	0.8931	0.035	25.464	0.000	0.824	0.962

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	-0.9655	-0.3579j	1.0297	-0.4435
AR.2	-0.9655	+0.3579j	1.0297	0.4435
AR.3	-6.9469	-0.0000j	6.9469	-0.5000
MA.1	-1.0006	-0.3441j	1.0581	-0.4473
MA.2	-1.0006	+0.3441j	1.0581	0.4473

Fig.16. The information of ARIMA(3,1,2) for MSFT

I then did a residual analysis, which evaluates the model from the perspective of residual. Based on the following four figures, the residuals of NVDA ARIMA model and MSFT ARIMA model concentrated on 0, and are respectively within 30 and 15. Most ACF of residuals are within the confidence interval. The QQ plots of two models are nearly linear.

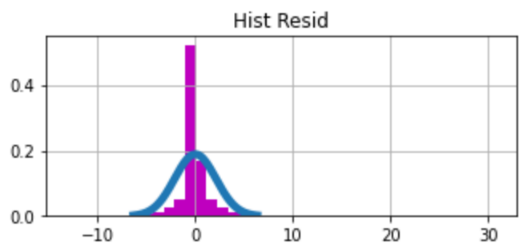

Fig.17. Residual distribution of NVDA ARIMA

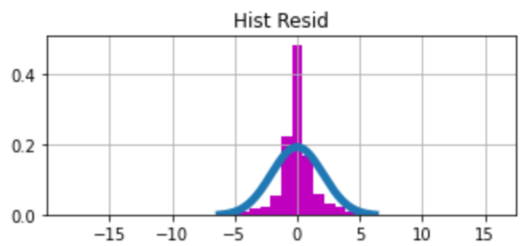

Fig.18. Residual distribution of MSFT ARIMA

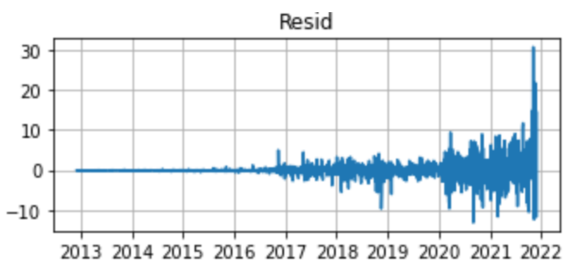

Fig.19. Residual of NVDA ARIMA

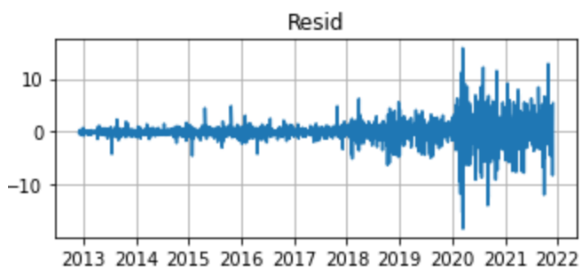

Fig.20. Residual of MSFT ARIMA

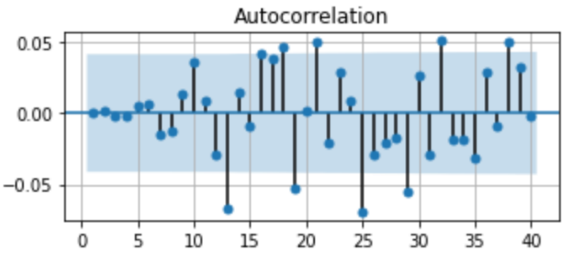

Fig.21. ACF plot of residual of NVDA ARIMA

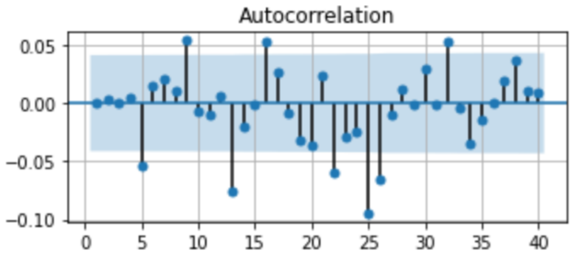

Fig.22. ACF plot of residual of MSFT ARIMA

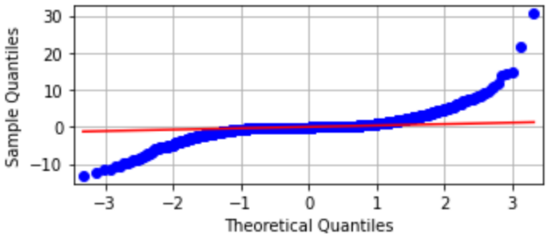

Fig.23. QQ plot of residual of NVDA ARIMA

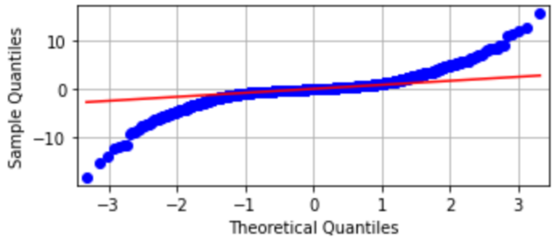

Fig.24. QQ plot of residual of MSFT ARIMA

In addition, I conducted a white noise test for the residual. The LB result for the two models are 0.8894 and 0.2220, which are greater than 0.05, so that the residual is white noise. Based on the four plots and LB test result, it is reasonable to conclude that the models adequately extract the information from the sequences.

4 LSTM Solution

LSTM stands for Long short-term memory. LSTM cells are used in recurrent neural networks that learn to predict the future from sequences of variable lengths. The main idea behind LSTM cells is to learn the important parts of the sequence seen so far and forget the less important ones realized by several gates, including input gate, forget gate, and output gate. Due to the LSTM's strength of modeling long sequence data, it is widely used in time series analysis.

4.1 Data Preparation

I first used min-max scale to conduct the data normalization, and split the data into train set (70%) with the size of 1701, validation set (20%) with the size of 443, and test set (10%) with the size of 192. The validation set is used for tune LSTM model, while the test set is the same as the test data for ARIMA model, which is used for the comparison of ARIMA and LSTM model in section 5. Then, I created a look back window with the window size of 60 to generate X, the features used as input for the model, and y, the expected outcomes.

4.2 Model Training

I used mean square error as loss function, Adam as the optimizer, set max epochs as 200 and batch size as 256. Furthermore, I applied dropout strategy and early stopping strategy to avoid overfitting.

Fixed Parameters

Loss	Mean Squared Error
Optimizer	Adam
Dropout rate	0.2
epochs	200
early stopping	10
batch size	256

Table 4. Fixed parameters used in LSTM model for NVDA and MSFT data

Then, I tuned the parameters of number of hidden layers (n_layer) and the node number of each layer (lstm_units), and tested the model's performance on validation set.

Tune Parameters		Validation set		
lstm_units	**n_layer**	**scaled_RMSE**	**unscaled_RMSE**	**unscaled_MAE**
20	**1**	0.0601	19.8864	11.4175
20	**2**	0.0477	15.7946	10.2786
20	**3**	0.1011	33.4478	21.7636
50	**1**	**0.0351**	**11.6040**	**7.9038**
50	**2**	0.0490	16.2164	10.4662
50	**3**	0.0752	24.8816	16.2985

Table 5. LSTM tuning results on NVDA data

Tune Parameters		Validation set		
lstm_units	**n_layer**	**scaled_RMSE**	**unscaled_RMSE**	**unscaled_MAE**
20	**1**	0.1020	32.2955	29.1651
20	**2**	0.0696	22.0354	16.8614
20	**3**	0.1098	34.7806	27.7567
50	**1**	**0.0224**	**7.0872**	**5.7430**
50	**2**	0.0285	9.0229	6.9937
50	**3**	0.0690	21.8637	16.9590

Table 6. LSTM tuning results on MSFT data

The parameters of the node number of each layer and the number of hidden layers for the best LSTM models for NVDA and MSFT are both (50, 1). The converge process of the best models is shown in fig.25 and fig.26. These two figures show that during training, the errors on train set are first decreasing and then remain nearly constant, which proves the convergence or both models. From fig. 27 and fig. 28, we can see that the loss of validation set is also first decreasing and then remain nearly unchanged. Therefore, the models are not overfitted. From fig.29 and fig.30, we can see that the true price lines and the predicted price lines are overlapped, which shows that the two LSTM models performs well.

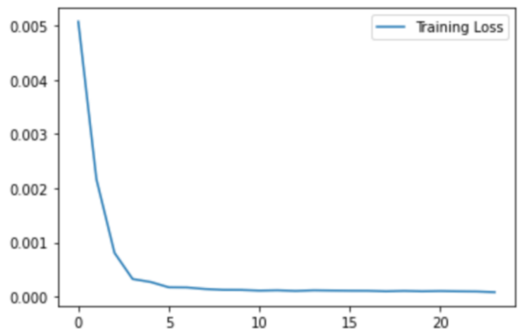

Fig.25. Training loss change during NVDA LSTM training

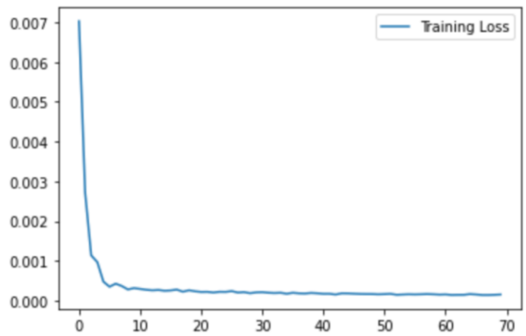

Fig.26. Training loss change during MSFT LSTM training

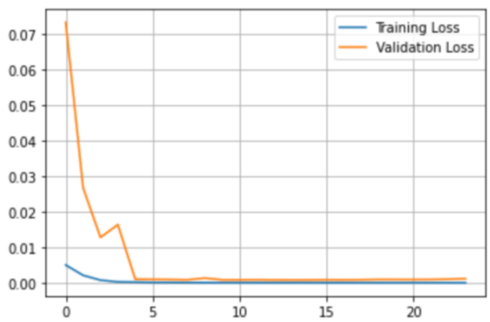

Fig.27. Loss changes during NVDA LSTM training

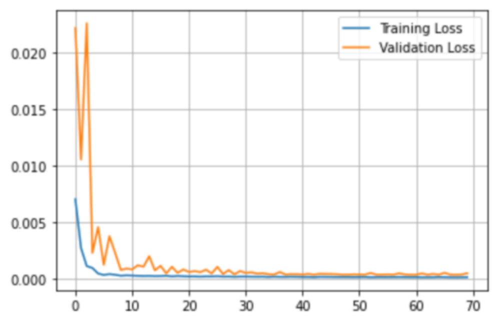

Fig.28. Loss changes during MSFT LSTM training

Fig.29. Predicting results of NVDA LSTM model

Fig.30. Predicting results of MSFT LSTM model

5 Discussion

Based on the best ARIMA and LSTM models for NVDA and MSFT price data, I made a comparison of their performance on the test set.

		ARIMA	LSTM
NVDIA	RMSE	178.68	15.43
	MAE	165.58	12.11
Microsoft	RMSE	85.82	9.40
	MAE	76.34	8.07

Table 7. Performance of ARIMA and LSTM model on test set

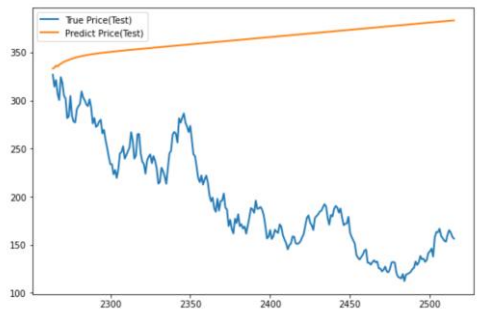

Fig.31. NVDA ARIMA performance on test set

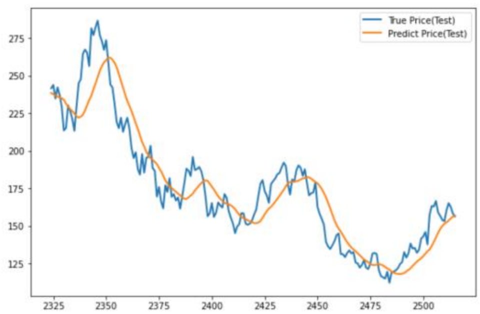

Fig.32. NVDA LSTM performance on test set

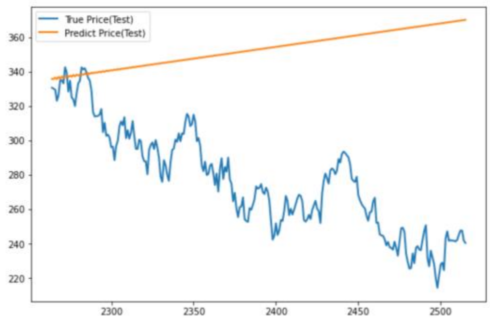

Fig.33. MSFT ARIMA performance on test set

Fig.34. MSFT LSTM performance on test set

From table 7 and fig31 to fig 34, we can see that and the LSTM model for Microsoft outperforms the LSTM model for NVDIA, and LSTM model can predict much better than ARIMA model. However, it is not rigorous to conclude that LSTM model is better than ARIMA model, because ARIMA model predicts the future stock prices only based on the trained model, which only contains the information of training data (previous stock prices), while although LSTM model is trained on the same training data, its prediction need input features, which are actually the stock prices of several days ago, so that during the prediction process, LSTM model also absorbs some information of test set. Therefore, I think ARIMA model is only good at predicting prices in s short term, while LSTM model is excel at making predictions in a long term.

6 Conclusion

To summarize, based on the stock prices of NVIDIA and Microsoft from 2017-12-1 to 2022-11-30, I built ARIMA model and LSTM model on both NVDA and MSFT dataset. For the best NVDA ARIMA mode, p and q should be 5 and 1. For the best MSFT ARIMA mode, p and q should be 3 and 2. The RMSE of best NVDA LSTM model is 15.43, the MAE of best NVDA LSTM model is 12.11; the RMSE of best MSFT LSTM model is 9.40, the MAE of best NVDA LSTM model is 8.07. However, because of the characteristics of ARIMA model's short term prediction, ARIMA models do not perform well on the test set.

Reference

Cai, Lier. (2017). Quantitative Investment Using Python.

Kutzkov, Konstantin. *"ARIMA vs Prophet vs LSTM for Time Series Prediction"*. Nov. 14, 2022. https://neptune.ai/blog/arima-vs-prophet-vs-lstm.

Zhang, Aston & Lipton, Zachary & Li, Mu & Smola, Alexander. (2021). Dive into Deep Learning.