

Stock Prediction of TSMC and Intel using Machine Learning

Ningjing Zhang*

Hwa Chong Institution (Junior College), Singapore, 269734

*Corresponding author: 182512h@student.hci.edu.sg

Abstract. The advancement of communications, computers, healthcare, military systems, transportation, renewable energy, and numerous more uses is made possible by semiconductors, which are a crucial part of electronic equipment. Semiconductors have been experiencing a shortage due to various reasons. Thus, the increasing demand for chips has attracted countless people to invest in the semiconductor industry. Intel Corporation and Taiwan Semiconductor Manufacturing Company Limited (TSMC) are two companies that can be said to be dominating the industry currently as they have been making very advanced chips. As two giant suppliers in the semiconductor industry, they are close substitutes and competitors. This essay is going to evaluate Intel and TSMC s' stocks for investors by using the method of machine learning to estimate and predict the future trend of the two stocks. The two models that will be used in the essay are the Linear model and the Long Short-Term Memory model (LSTM), showing the outcome of prediction and determining which stock is better for the investment.

Keywords: Semiconductor; stock prediction; machine learning; LSTM model; linear model.

1. Introduction

The Stock exchange is the mediator that allows buying and selling of shares [1]. The act of attempting to anticipate the future value of a business stock or other financial instrument traded on an exchange is known as a stock market prediction. A stock's accurate future price prediction could result in a sizable profit [2].

The growing number of electronic devices, data centers, home appliances, and cars all require semiconductors. Additionally, the continuation of the pandemic has caused great changes in people's work and lifestyle. Home office and home vacations have become the habitual routine of most people. Teleconferencing and online entertainment require a large amount of electronic equipment, which increases the demand for electronic consumer goods, and correspondingly, the demand for chips continues to rise [3]. The semiconductor market is currently in a phase of active development and growth. Today's market is over \$50 billion and its growth potential is in the trillions of dollars over the next 10 to 20 years [4]. Moreover, semiconductors are essential to the economy. Semiconductors may contribute only a small percentage of global GDP, but they fuel trillions of dollars of goods and processes. In addition to growing global demand, demand for semiconductors is also strongly driven by digital transformation and further accelerated by the pandemic. The semiconductor industry association reported that global semiconductor sales rose 6.6 percent to \$440 billion in 2020, even as global GDP shrank 3.5 percent. This scale and value mean that the role that semiconductors play in global products and operations or in the broader economy is important to leaders in all industries [5]. Therefore, the semiconductor industry becomes a good choice for people to invest in.

Intel and TSMC, with their advanced technology and efficient production capacity, have won the favor of many well-known customers. Because chip technology is a core force of scientific and technological development, as long as the skill is mastered, one can stand in the semiconductor industry. Intel is the world's second-largest semiconductor company and the first company to launch a central processing unit (CPU) based on the x86 architecture, headquartered in California, USA [6]. When it was launched in 1987, TSMC invented the semiconductor Dedicated IC Foundry business model. TSMC produced more than 12,300 products for a variety of applications across a range of end markets, including smartphones, high-performance computing, the Internet of Things (IoT), automobiles, and digital consumer electronics, while serving around 535 customers [7]. Taiwan Semiconductor Manufacturing is a true industry behemoth with a well-known reputation for chip

manufacturing. It has continuously produced outstanding operating results and attracted the interest of investors, such as Warren Buffett. TSMC is the top chip manufacturer in the world, and its current pricing is enticing to long-term investors [8]. TSMC and Intel have the best technology in the industry and are trying to make chips even thinner. They are competitors and both dominate the industry, attracting large amounts of investment and supplying raw materials for many products.

Software systems may anticipate outcomes more correctly without explicit instructions by using machine learning (ML), a kind of artificial intelligence (AI). Machine learning algorithms use previous data as input to predict new output values. Machine learning is important since it helps with the creation of new products and gives companies and investors an overview of consumer behavior trends and company operating patterns [9]. In this essay, the Long Short-Term Memory model (LSTM) and Linear model will be used to evaluate the pattern of the stocks for Intel and TSMC, doing a prediction of the future trend of the stocks, and showing investors which stock should be chosen to invest.

The LSTM model is used in the field of Deep Learning. It is a variety of recurrent neural networks (RNNs) that may pick up long-term relationships, particularly in situations involving sequence prediction. [10]. The Linear model describes a continuous response variable as a function of one or more predictor variables. Linear models are helpful for understanding and predicting the behavior of data. Linear regression is a statistical method used to create a linear model [11].

By using the database of the two stocks from Kaggle, the LSTM and Linear models will be built to show the trend of the stocks' future. First, train data will be split from test data to do a simulation of the prediction to see if the model is suitable, and also Mean Square Error test will be used to determine the accuracy. Later, the essay will decide to use one of the models to make future predictions.

2. Methodology

2.1 Visualize the Data

2.1.1 Database

By downloading the database of Intel from Kaggle [12], and running the code to check the datasets, the information about Intel's stock can be seen including the stock information about Open price, Close price, Adjusted Close price, High, Low and Volume from 1980 17th of March to 2022 20th of May. The database of TSMC has the same content but only from 2009 31st of December to 2022 2nd of September [13]. Comparing the first value and the last value recorded in the stock data, a huge difference can be seen, so it can be concluded that over a period of several decades, although reasons such as inflation should also be taken into account, the semiconductor industry is growing extremely fast.

2.1.2 Intel trend

The trend of the stock for Intel is plotted using Matplotlib. (Shown in Fig.1) The graph shows that there are fluctuations in Intel's stock. The stock had a sharp increase in 2020 January and 2021 April. However, the stock experienced quite a few massive drops. The final part of Intel's graph shows a trend of decrease.

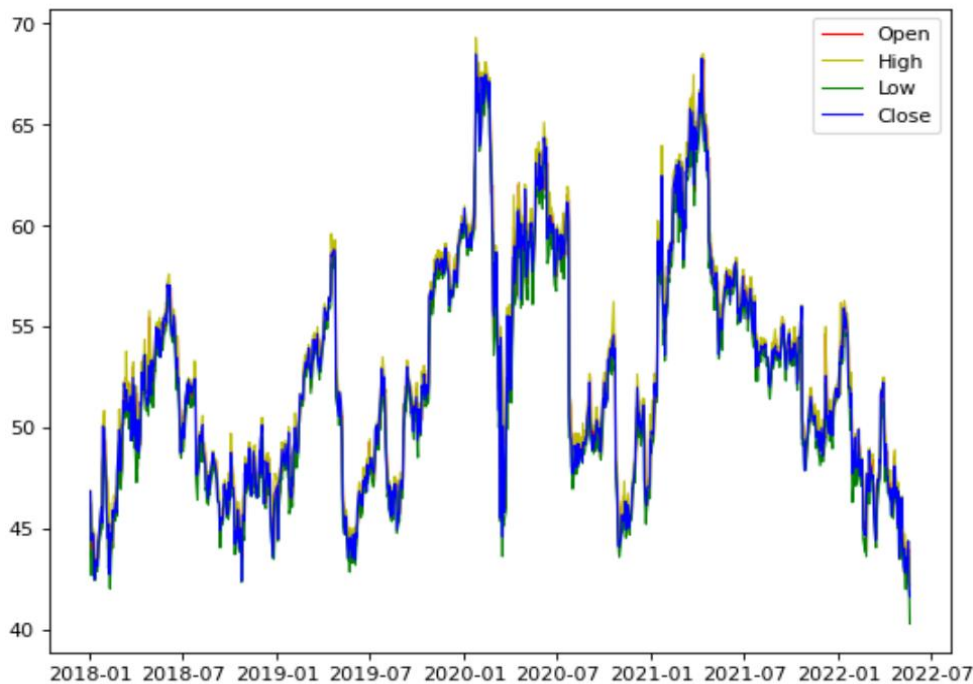


Figure 1. Intel's Stock Trend

2.1.3 TSMC trend

TSMC's rapid growth can be seen clearly. The company has reached a considerable level in a very short time. Very large plots dominate the industry with more advanced technology. This also reflects the potential of TSMC. The trend of the stock for TSMC is plotted using Matplotlib. (Shown in Fig.2) The graph shows that there is a very obvious trend of increase, especially from 2020 to 2021, the growth is remarkable and even though there is a slight decrease during 2021 and 2022, the stock of TSMC has not changed very much like Intel's stock.

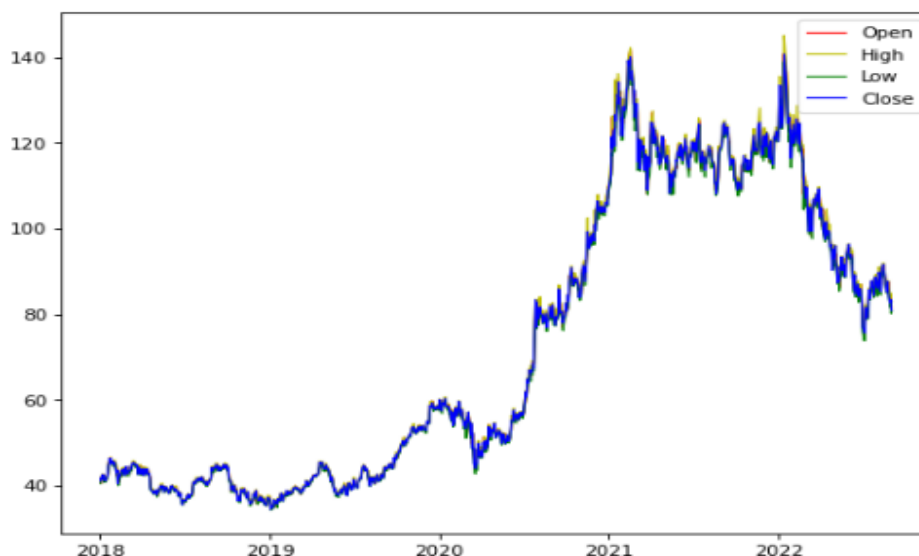


Figure 2. TSMC's Stock Trend

2.2 Evaluation of Dataset

Comparing Intel with TSMC's stock, TSMC seems to have a steadier development, while Intel's stock fluctuates a lot, which means investors may face more risk. However, both stocks are experiencing a downward trend in 2022. Intel's stock is down because the technology sector in the U.S. stock market has suffered a severe correction this year. Whereas TSMC's stock may be down

because of the US trend, in addition to the impact of chip law 2022. The United States invited TSMC to build factories in the United States and promised that the government would subsidize semiconductor industry chain manufacturers such as Taiwan Semiconductor Manufacturing, Samsung and Intel with huge subsidies for plant construction, hoping that TSMC could assist the United States to rebuild the manufacturing industry chain of chip manufacturers. However, the lack of subsidies resulted in the loss of 300 billion dollars of TSMC's market value [14]. Another reason is the rise of the mainland's chip industry. Although China has the largest chip consumption market, it is still some distance from the most advanced level except for the design link, but fortunately, the whole chip industry chain has a layout, and the negative impact of sanctions and repression in the short term is not small, but also brings opportunities for domestic substitution, becoming less dependent on TSMC [15].

3. Results and Discussion

3.1 Linear Model

To complete the stock prediction in a machine-learning manner, this paper presents two selected models: the Linear model and Long Short-Term Memory Model (LSTM). Since the semiconductor themselves exist as a technological product that especially keeps pace with the times, recent data are used from the datasets. The content from 2018 1st January to 2022 19th May is taken out for working to ensure the prediction outcome is suitable to be taken as a reference to decide which stock is a better choice for people to invest in. The link between the two elements or variables can be predicted using a linear regression model.

3.1.1 Prediction of test data

Close price values are used. The last 300 data are split for tests and the rest in the front are used for training. The Linear regression model is imported.

Program predicting the test data is written, and the graphs for linear models are plotted. (Shown Fig.3 and Fig.4)

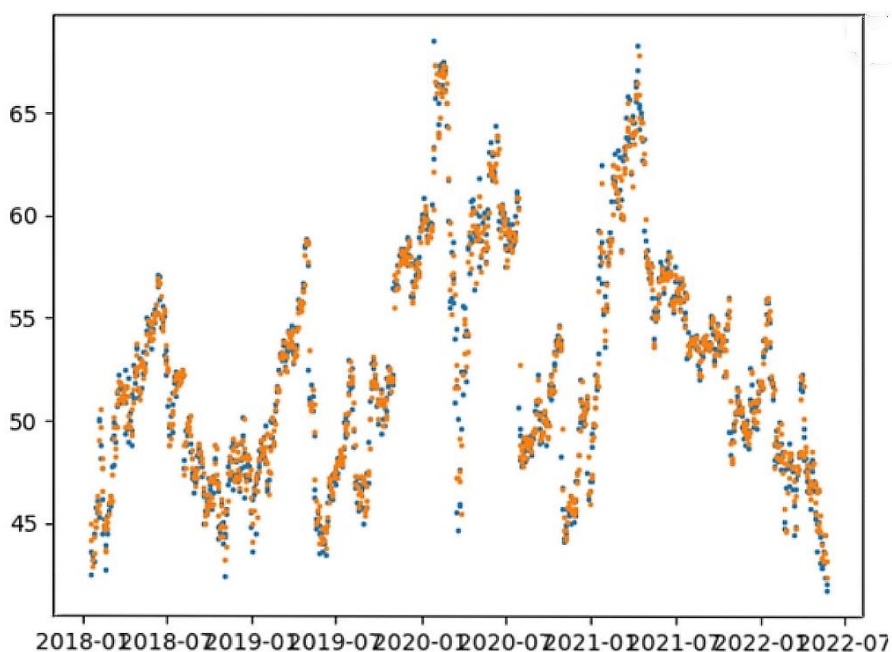


Figure 3. Linear model Intel Test Data Prediction

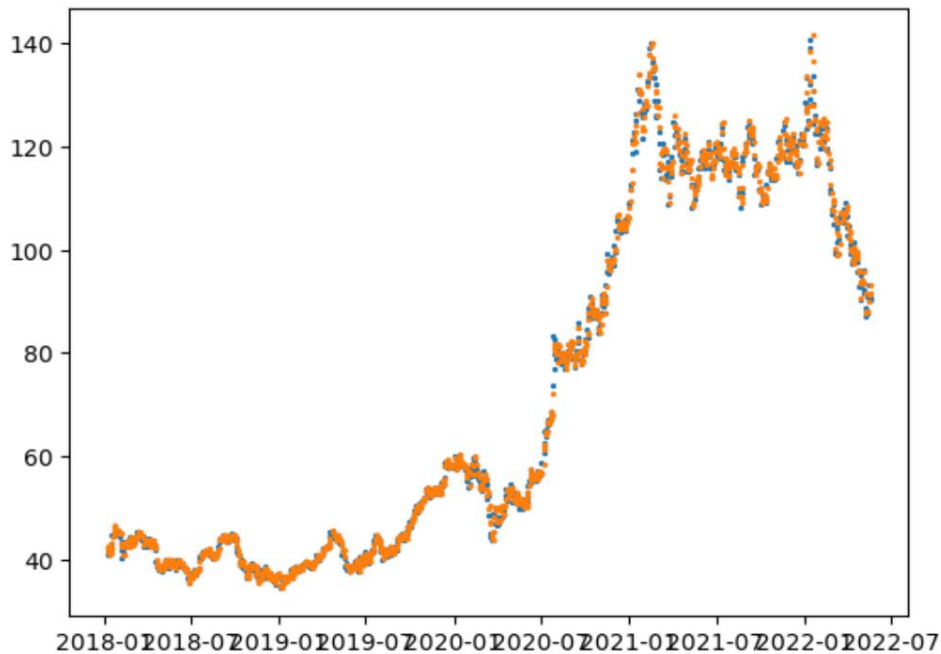


Figure 4. Linear Model TSMC Test Data Prediction

The Linear model was trained to use the data from the previous day to predict the close price for the next day to ensure the accuracy of the prediction. It can be seen from the chart that the overlap of points of different colors is still very high. The blue point is the actual value of the close price in the original database, while the orange point is the result of the linear model prediction. It can be seen from the Intel chart that there is still some deviation, especially in places where the stock price fluctuates greatly. On the TSMC prediction chart, the test data and the actual data are very consistent. The Linear model is one of the simplest models in Machine Learning. If more days are used for the prediction, the values predicted for the last 300 data will not match the test data so much. The model currently used can predict the future of the stock correctly in the very short term.

3.1.2 Accuracy test

From the above graphs, since the characteristic of the linear model is that it is only based on one day's data to make predictions, it fits the original data very well, and the degree of accuracy cannot be determined from the chart. Therefore, to do the model comparison, a mean square error test is used (MSE). MSE calculates the mean or average of the square of the difference between actual and estimated values. The formula of the mean square error test is $\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$, where 'n' represents the number of data points, Y_i represents observed values and \hat{Y}_i represents the predicted values. After running the program to calculate MSE, the result is very small. For Intel, the MSE value is 1.4857 and for TSMC, it is 3.2971.

3.2 LSTM Model

LSTM is a type of recurrent neural network which can successfully train very large architectures and it is very suitable for solving time series prediction problems.

3.2.1 Prediction of test data

Test data used are the last 300 for Close Price value, train data are the rest in front, LSTM model is imported to make the prediction. The LSTM graphs for both stock's test data predictions are plotted. (Shown in Fig.5 and Fig.6)

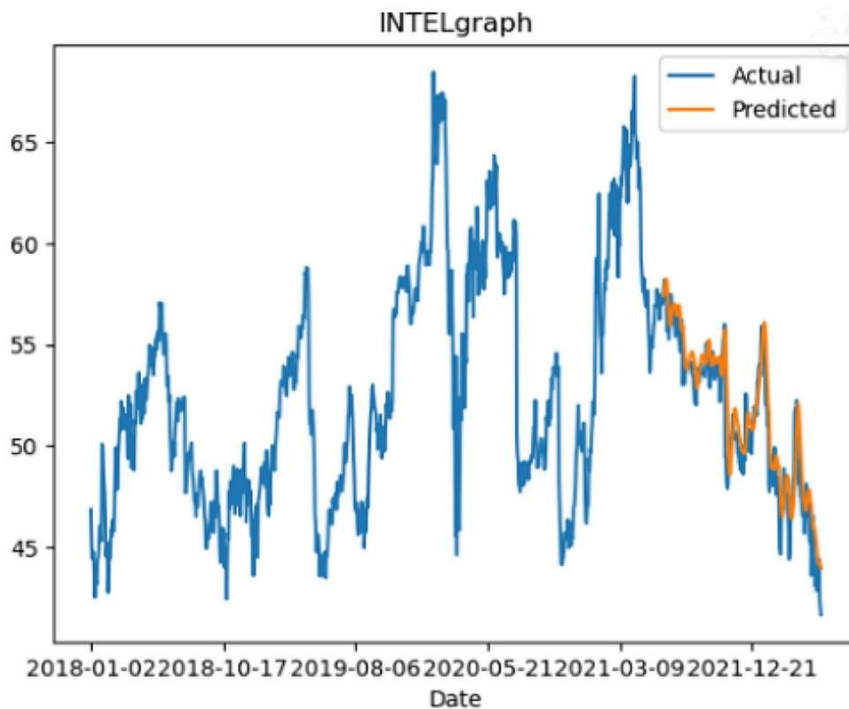


Figure 5. LSTM Intel Test Data Prediction

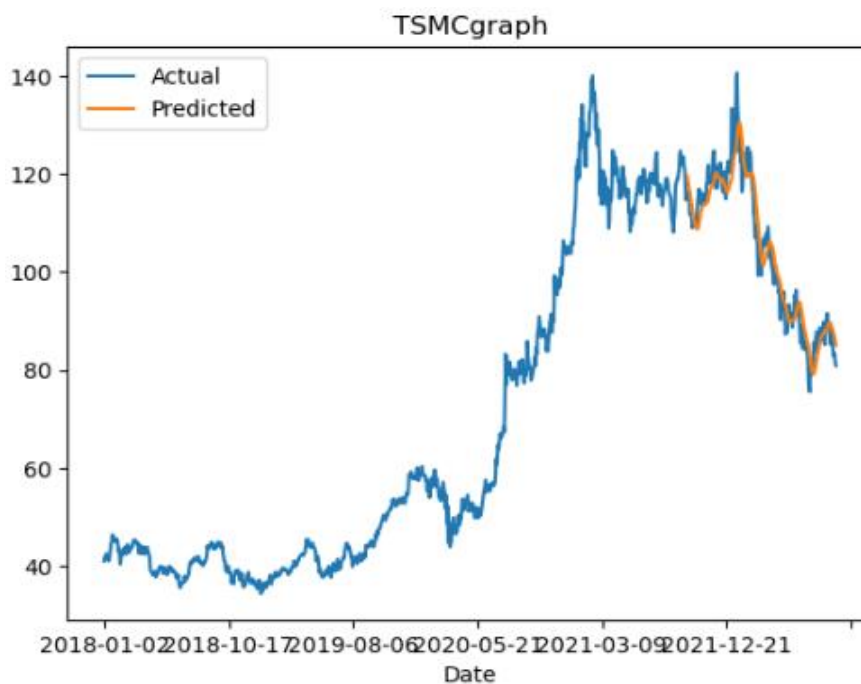


Figure 6. LSTM TSMC Test Data Prediction

The LSTM model itself is a type of relatively complex deep learning. It is a very accurate model, which is specially used to solve problems such as predicting stocks. The blue line in the image is the data of the original database, and the orange line is the prediction of the last three hundred test data. The predicted line clearly and pertinently depicts the trend of the stock market, and overlaps with the graph of the original data, which is enough to show the accuracy of the model.

3.2.2 Accuracy test

The MSE test is used to calculate the accuracy of the model. In such a way, LSTM and Linear models can be compared. The Mean Square Error Test calculates the accuracy. The result for Intel is 2.0889, and for TSMC is 16.7361.

3.3 Future Prediction

Even though the model is more accurate if MSE is closer to 0 which shows that Linear model is better, the MSE values of both the linear model and the LSTM model are very low. Therefore, showing both methods are suitable and accurate for doing future forecasts. Since both models can be used for future prediction, LSTM will be used in this essay to show the future trend. The future sequence for both of the graphs is plotted. (Shown in Fig.7 and Fig.8)

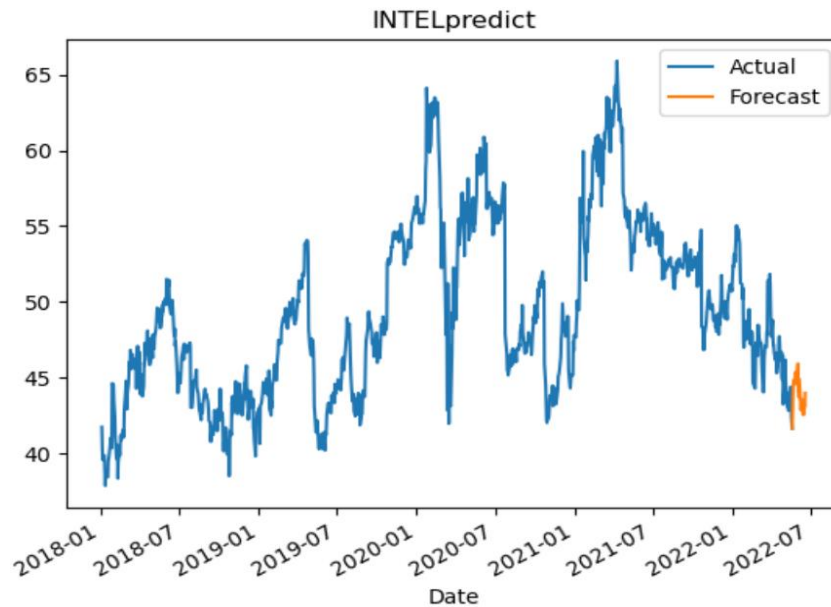


Figure 7. LSTM Intel Future Prediction

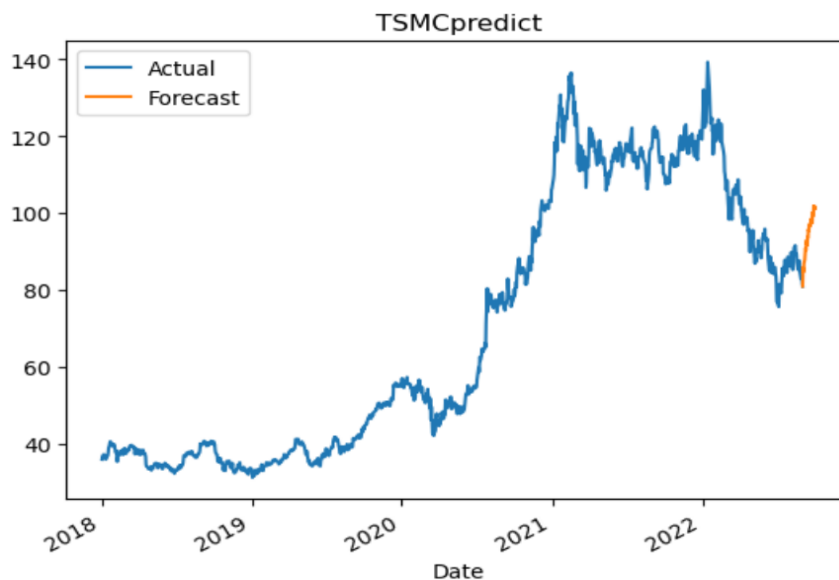


Figure 8. LSTM TSMC Future Prediction

The LSTM model helps to predict the close price for the next 30 days and plotted the trend on the graph. The blue line shows the original stock trend by using the values from the database and the orange line in the graph shows the future sequence. For TSMC, it is increasing. The line is steep as it has a large gradient. This shows that TSMC's close price will have a large increase in the future, and it is also a relatively stable investment target since the line plotted can be seen as a straight line because there are no curls or kinks. This implies that TSMC can give investors a good rate of return in the next month. For Intel, similar to its previous changes, the orange line has many twists and turns, which proves that there will be fluctuations in the future close price. This shows the instability and

investment risk of this stock. The main trend is declining. Although there is a small increase in the end, Intel will bring investors a certain degree of loss.

4. Conclusion

The LSTM is a very suitable model to evaluate time series data by learning from past stock market data and finding patterns to make future predictions. The drawback is it takes quite a long time for the code to run compared with other models. The conclusion is that according to the forecast, TSMC's stock is trending up while Intel's stock is trending down. So, in June, the next 30 days, TSMC would be a better investment choice than Intel. Additionally, by reading the articles and news, TSMC is working on some more advanced technology showing it was ahead of the game, thus making TSMC a better choice for investment. However, the stock market is highly variable and the stock price is affected by numerous factors. Moreover, the problem is that the label data used in the prediction is invariably very small and this kind of short-term prediction will not be very effective when it is traded in the market. In the future, the data for a longer time should be used to do the prediction even though it becomes less accurate, but this way of prediction will be more effective and useful to be used as a reference. Therefore, the results predicted by LSTM can only be used as a reference which cannot accurately calculate the stock price.

References

- [1] Avijet Biswal. Stock Price Prediction Using Machine Learning: An Easy Guide. November 15 2022. Retrieved on December 29, 2022. Retrieved from <https://www.simplilearn.com/tutorials/machine-learning-tutorial/stock-price-prediction-using-machine-learning>.
- [2] Shoanmike, Wikipedia. Stock Market Prediction. December 24 2006. Retrieved on December 29, 2022. Retrieved from: https://en.m.wikipedia.org/wiki/Stock_market_prediction.
- [3] TianTian Fund Network. Why is the chip stock so popular? The reason was found. June 18, 2021. Retrieved on December 29 2022. Retrieved from https://mbd.baidu.com/newspage/data/landingsuper?rs=3754668000&ruk=xIDGehzS4KqSIRAICQrFJw&isBdboxFrom=1&pageType=1&urlext=%7B%22cuid%22%3A%22_Pvpa0aFSfjOuHfqYav8i_auvi0_iS8C0aS58_aEB8KX0qqSB%22%7D&context=%7B%22nid%22%3A%22news_9516693817532852399%22%7D.
- [4] Market Trend. Is it Worth Investing in Semiconductor Stocks? August 8 2022. Retrieved on December 29 2022. Retrieved from: <https://www.analyticsinsight.net/is-it-worth-investing-in-semiconductor-stocks/>.
- [5] Parichart Jiravachara. How semiconductor plays as a crucial element for economy. March 2022. Retrieved on December 29 2022. Retrieved from: <https://www2.deloitte.com/th/en/pages/about-deloitte/articles/semiconductor-eng.html>.
- [6] Wikipedia. Intel. February 11 2008. Retrieved on December 29 2022. Retrieved from <https://zh.m.wikipedia.org/zh-cn/%E8%8B%B1%E7%89%B9%E5%B0%94>.
- [7] Taiwan Semiconductor Manufacturing Company Limited. "About TSMC". Retrieved on December 29 2022. Retrieved from: <https://www.tsmc.com/english/aboutTSMC>.
- [8] Retirement Dash. Taiwan Semiconductor (NYSE: TSM): One of the Best Bets in Tech. December 12 2022. Retrieved on December 29 2022. Retrieved from: <https://www.tipranks.com/news/article/why-tsm-stock-is-one-of-the-best-bets-in-the-tech-space>.
- [9] Ed Burns. Machine Learning. November 14 2019. Retrieved on December 29 2022, Retrieved from: <https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML>.
- [10] Intellipaat. What is LSTM? Introduction to Long Short-Term Memory. December. 13, 2022. Retrieved on December 29 2022. Retrieved from: <https://intellipaat.com/blog/what-is-lstm/#:~:text=%20LSTM%20stands%20for%20long%20short,especially%20in%20sequence%20prediction%20problems>.
- [11] MathWorks. "Linear Model". Retrieved on December 29 2022. Retrieved from: <https://www.mathworks.com/discovery/linear-model.html>.

- [12] Verma, Arpit. "Intel Stock Data." Kaggle, March 25, 2022. Retrieved on December 29 2022. Retrieved from: <https://www.kaggle.com/datasets/varpit94/intel-stock-data>.
- [13] ProgrammerRDAI. TSMC Stock. Kaggle, May 23, 2022. Retrieved on December 29 2022. Retrieved from: <https://www.kaggle.com/datasets/ranugadisansagamage/tsmc-stock>.
- [14] LuoFeiYu. TSMC did not expect! After 52 billion 'disappears', the market value evaporated by 300 billion US dollars! July 28 2022. Retrieved on December 29 2022. Retrieved from: https://ml.mbd.baidu.com/r/RRJITaDO0M?f=cp&rs=2958965577&ruk=xIDGehzS4KqSIRAlCQrFJw&u=4e0bc1bd87676377&urlext=%7B%22cuid%22%3A%22_Pvpa0aFSfjOuHfqYav8i_auvi0_is8C0aS58_aEB8KX0qqSB%22%7D.
- [15] Market Finance. It was a bit unexpected, TSMC's stock price has fallen 59%. October 24 2022. Retrieved on December 29 2022. Retrieved from: https://mbd.baidu.com/newspage/data/landingsuper?nid=news_9037812780712480511&_refluxos=&pageType=1&wfr=.