ETF Prediction of Leading Southeast Asian Countries Using Different Machine Learning

Weiyi Mu[1,†] Zihan Nan [2,†] Zhouhang Ren [3,†] Zhixin Ye [4,†,*]

[1] *Department of Engineering, McMaster University, Hamilton ON, L8S 3L8, Canada*
[2] *Department of Economics and Management, Beijing Jiaotong University, 100044, China*
[3] *Department of Applied Economics, Macau University of Science and Technology, 999078, China*
[4] *School of Mathematics and Statistics, University of Glasgow, Glasgow, G12 8QQ, United Kingdom*
[*] *Corresponding author. Email: 2700300Y@student.gla.ac.uk*
[†] *These authors contributed equally*

ABSTRACT

The current health crisis plays a significant role in the stock market. This study aims to investigate the impact of COVID-19 on the Southeast Asia stock market, especially in Singapore, Thailand, India, Indonesia, Malaysia, and the Philippines. For this purpose, this study considered the influence on the Exchange Traded Fund (ETF) from the date the first COVID-19 case was reported in each country and the lookback period. The collected data covered the period between 3 February 2012 and 18 March 2022. Using the method of Long-Short Term Memory RNN (LTSM) to predict ETF trading with three different levels of lookback parameters of 60, 30, and 15. In terms of Singapore and India, 60 days lookback parameters had the best performance for the whole prediction. For the Philippines and Thailand, 60 days lookback parameters predicted the best before the first COVID-19 case was confirmed in each country and 15 days lookback parameters had the best prediction during the COVID-19 period. The results illustrated that most of the six countries mentioned in this study showed that with the increase of the lookback parameters, the model predicted more accurate; however, for the individual country, the lookback parameters had some differences due to the historical stock price and the COVID-19 situation in each country.

Keywords: *Machine learning, Southeast Asia, stock market, LSTM.*

1. INTRODUCTION

1.1. Background

The spread of COVID-19 has evolved from an unexpected pneumonia virus transmission in a Chinese city to a global pandemic, causing serious damage not only to human lives, but also to the financial markets. COVID-19 led to economic turmoil, business closures, global supply chains chaos, and millions of people were blockaded. Since 2020, several international firms are involved in divestment due to massive panic and market volatility. To alleviate this situation, several countries have introduced different stimulus measures, including lowering the interest rate and direct monetary support. In a short time, the Global economy did move towards recovery and the stocks rose continuously, but the global supply chain was slow to recover due to the back-and-forth of the epidemic. So generally, the COVID-19 still harmed stock markets, especially in Asia and Europe.

A large number of studies have focused on some emerging stock markets like Latin America and Europe, however, relatively few articles have focused on the South Asian market, even though India and some countries have lowered the international trade barrier since the economic transformation. Because of the growing interest of the global investors in the emerging markets especially the South Asian stock market, the stock markets of South Asian countries have outperformed the developed countries recently. These economies are rapidly globalizing by deregulating the financial system these years, increasing the market capitalization of South Asian stock markets.

South Asian's stock market has been volatile in recent years. Also, the emergence of the COVID-19 pandemic causes a global economic slow-down and fluctuation in stock prices. With no vaccine or targeted therapy for COVID-19 available to date, this uncertainty of economic and financial distress is not expected to end any time soon. It makes predicting future stock prices and

Y. Jiang et al. (Eds.): ICEDBC 2022, AEBMR 225, pp. 670–676, 2022.
https://doi.org/10.2991/978-94-6463-036-7_100

their trends more difficult. To get an accurate prediction result, a long-and-short-term memory (LSTM) model is designed to be applied to six South Asian countries which are Singapore, India, Malaysia, Philippines, Thailand, and Indonesia.

1.2. Related research

Brorvkova and Tsiamas used the LSTM neural networks to predict stock price. They used an ensemble model with plenty of LSTM neural networks to build a framework. And then used this model to predict 44 US large-cap stocks and compared the results with two other predict models: the lasso and ridge logistic regressions. They found that the LSTM neural networks model could provide better results [1]. Lu et al. researched to verify the advantage of LSTM neural networks in stock prediction. They used the CNN-LSTM, RNN, MLP, CNN, RNN, LSTM, and CNN-RNN models to predict stock price. The data they used was the stock prices from July 1 1991 to August 31 2020. By comparing the results of these models, they found that the CNN-LSTM model was the best one because it had the smallest MAE and RASE [2]. Sen et al. proposed a deep learning-based regression model that was built on a LSTM network that automatically searched the network, extracting historical stock prices for a specified pair of start and end dates based on the stock's ticker name and forecasting the future stock prices. Moreover, the predicted values of the stock prices were used as the basis for investment decisions, and returned on the investment calculated. The results enabled them to compare the profitability of the sectors from the perspective of stock market investors [3]. Moghar and Hamiche used the ML algorithm which was based on LSTM RNN to get the future adjusted closing prices for a portfolio of assets, aiming to figure out the best algorithm which could obtain the future prediction of the portfolio. After training the target algorithm on two different assets, they found that the model was confident enough to trace the change in the open prices [4]. Sunny et al. put forward a new framework about the price of stock based on the LSTM model and Bi-Directional Long Short Term Memory (BI-LSTM) model. By using Deep Learning algorithms such as this framework with hyperparameter tuning, the prediction could be generated by the highest level of accuracy with a low level of RMSE [5]. Pawar et al. worked on predicting the stock price by using the RNN and LSTM, considering the influence from buyers or customers in the market. Several other machine learning algorithms were also compared during the research. After assessing the models, the RNN-LSTM model was evaluated as the model with higher accuracy. In addition, this final model was suitable for both individual buyers and corporate customers [6].

According to the four countries' markets in Southeast Asia, including Indonesia, the Philippines, Cambodia, and Myanmar, Wyman found that financial inclusion could be impacted in a positive way by the digital financial services in business models. Wyman mentioned only a narrower group of other new markets could be impacted in the same way due to the different stages in different countries [7]. Narayan et al. found that the stock in Southeast Asia was becoming more integrated, due to the analysis of the relationship between Southeast Asia stock markets after the 1990s stock markets opening. In addition, the time-varying parametric model proved that the Philippines, and Thailand stock market results had a stronger relationship with Singapore [8]. According to the autoregressive conditional heteroscedasticity in five Southeast Asian countries (Indonesia, Malaysia, Singapore, Philippines, and Thailand) between 2001 and 2015, Wahyudi et al. used time series to analyze how the inflation, GDP, and five other macroeconomic variables affect the composite index in a positive way or a negative way [9].

Mehtab et al. used different deep learning models to predict the stock price of NIFTY 50 index values on the Indian national stock exchange. They built 8 models and tested them. By comparing the product-moment correlation coefficient and RMSE of these models, they found that the LSTM deep learning models provided the best result [10]. Cooray and Wickremasinghe examined the efficiency in the stock markets of India, Sri Lanka, Pakistan, and Bangladesh. They used Dicky-Fuller Generalized Least Square and Elliot-Rothenberg-Stock to study the efficiency of weak forms of the stock market. It turned out that Bangladesh was not supported by DFGLS and ERS tests strongly, so they used Cointegration and Granger, and the results showed a high degree of interdependence among South Asian stock markets [11]. Kumar and Dhankar used the Granger causality test in short-run causal relations, which highlighted the significant autocorrelations in stock returns. The result of the Granger test indicates that the "F" statistics of India, Sri Lanka, and Pakistan are significant at the five percent level of significance at 2 and 5 lags. While the Bangladesh stock market is found to have no shorts-run causality to DJGI. So, because it did not respond to international market volatility in the short term, it may be a good destination for short-term investment. GARCH class models are also used to test the long-run spillover effects of stock returns. Results highlighted the significant co-integration relationship between South Asian stock markets and international markets. Investors with long investment horizons would not benefit from the portfolio [12].

1.3. Objective

The main objective is to predict the value of the Exchange Traded Fund of Southeast Asia from the previous data using machine learning. In the second chapter, the methodology which would be used to learn the pattern of the data, as well as the data used in the

research, will be introduced. The main finding of the research such as the basic results of the predictions produced by the method and the difference between the true value and the predictions will be presented in chapter 3, while chapter 4 will give the discussion based on the outcome and analysis related to the research background.

2. METHOD

2.1. Methodology

Deep learning is a branch of machine learning, a new way of learning from data. Deep learning emphasizes learning successive layers, hierarchies of cascades, many non-linear units each layer used for feature extraction and transformation, the next layer of input is the output of the previous layer, and each deeper layer represents a deeper abstraction of data. Modern deep learning typically consists of dozens or even hundreds of successive presentation layers, all of which are learned automatically from training data.

Recurrent Natural Network (RNN), a class of neural networks for processing time series data, is a common neural network structure that has been successfully applied to a wide range of problems such as Neuro-Linguistic Programming (NLP), speech recognition, and machine translation. The RNN is a commonly used neural network structure for processing time series data.

2.1.1. Basic structure

The recurrent neural network mainly consists of an input layer, an implicit layer, and an output layer, which are similar to traditional neural networks, although the self-looping W is indeed one of its special features, and its structure is as follows.

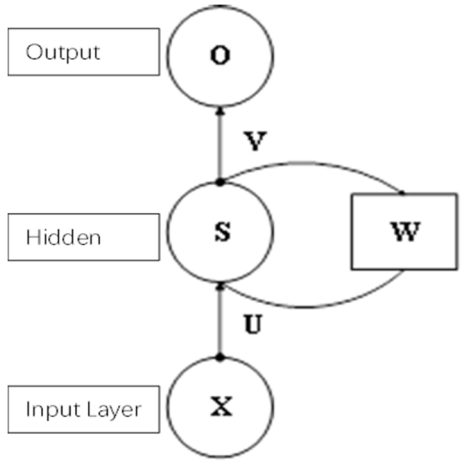

Figure 1 Structure of a recurrent neural network

X is the input variable, U is the input matrix of weight from the hidden layer, matrix W is the weight from the country to the hidden layer, S and V are the result output of the weight matrix from the hidden layer to the output layer and O. It is easy to see from the figure above that it shares parameters in such a way that the corresponding W, U, and V for each time node are constant. This method not only realizes parameter sharing, but also greatly reduces the number of parameters.

2.1.2. Long-and-short-term memory networks

Long-and-short-term memory networks are a special type of RNN that is capable of learning long time dependencies. They were proposed by Hochreiter & Schmidhuber [13] and have been improved and generalized by many others.

LSTM uses two gates to control the contents of cell state C: a forgetting gate, which determines how much cell state C_{t-1} is stored in the current state C_t from the previous moment; an input gate, which determines how much input X_t from the network is stored in the cell state C_t at the current time. The LSTM uses an output gate to control how much of the cell state C_t is output to the CURRENT output value h_t of the LSTM. The loop structure of LSTM is as follows:

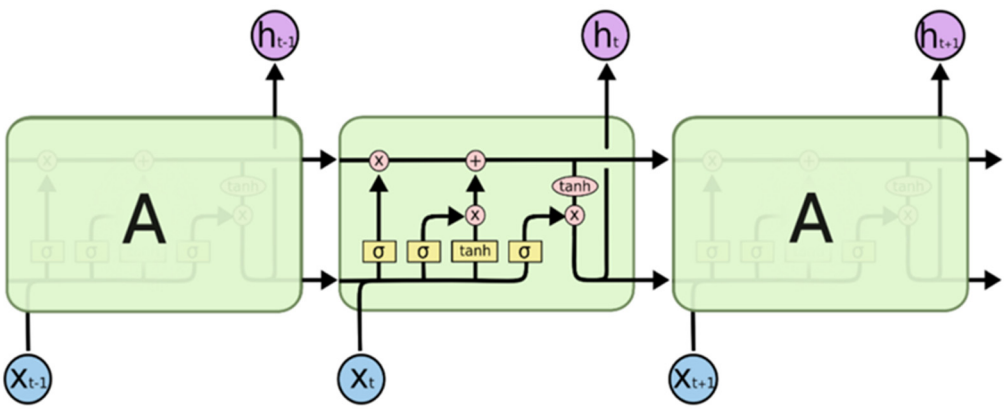

Figure 2 Structure of the LSTM recurrent neural network

Where the black lines with arrows indicate the direction of data flow, the input data is in vector form, the pink circle indicates the vector operation done on the two data, the yellow box indicates the mapping to be done on the data vector, and the black line divided into two indicates that the data vector has two uses.

2.2. Data processing

The data this time comes from Yahoo Finance which is from 2012.2.3 to 2022.3.18. After standardizing the data, it is clear that the date is recorded according to the year, month, day, hour, and minutes. By merging it into a time value, it was set into the index of a Data Frame. In the import data, the time is a categorical variable, while the other items are numeric variables.

Predicting future trend conditions from historical trend data is a problem that LSTM neural networks are better at handling. When dealing with time series, data from past periods (t-n to t-1) is usually used to predict current data (t). The data is therefore collated in this way and a dataset is generated for training the model. Here, only the concentration of the value of ETF needs to be predicted.

This time, we mainly make predictions through a 2-layer neural network, in which we use the results predicted by the first layer as the input of the second layer, and finally, predict the results. In this case, due to the limitation of the amount of data, using all the neurons will not only waste time, but also waste energy. Therefore, first of all, we will use some neurons to ensure accuracy and speed. Here we mainly use the parameters in KREAS to set the sample is determined to be extracted. This time we set it to 0.2, so only 20% of the samples participate in the modeling. We use ADAM as the model optimization algorithm, and the model optimization algorithm is accurate.

3. RESULT

In general, the predictions obtained by the larger lookback parameters will be closer to the actual situation than the predictions obtained by the smaller lookback parameters.

For Singapore, the 60-day lookback predictor curve (Figure 3) was the closest to the true value, while the 30-day lookback and 15-day lookback forecast curves also depicted the general trend of the actual situation, with some deviations in the value: the forecast value obtained by the 30-day lookback was higher than the true value, while the forecast values obtained by the 15-day lookback was lower before the outbreak of the Covid-19 and the fluctuations after the outbreak of the Covid-19 was smaller than the true value. The RMSE of these models also illustrated the same result that the model with 60-day lookback has the lowest value (0.929) of RMSE generally. The model with 15-day lookback gave the most accurate prediction of the downward trend of Singapore's ETF value after the outbreak of the Covid-19 (RMSE=1.047), but the resulting values were still significantly different from the real situation.

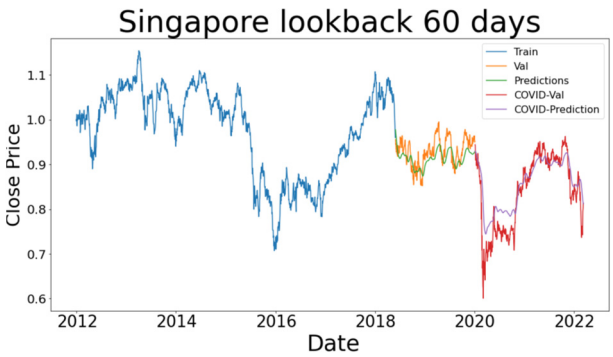

Figure 3 Prediction of Singapore using 60-day lookback

The results obtained by applying the model to the Indian stock market dataset are more intuitive. The value of ETF in India has been more volatile in recent years, and the increase in the curve has been more pronounced. The prediction curve obtained by the 60-day (Figure 4) lookback was basically the same as the real curve (RMSE=2.058), and as the value of the number of days to be looked back gradually decreases, the prediction curve becomes flatter and flatter, and the degree of deviation from the real curve is also getting larger and larger. It is worth mentioning that the prediction obtained by the 30 days lookback and 15 days lookback. From the figure, it is easy to figure out that the models with 30-day lookback and 15-day lookback performed better on the prediction of the decline in ETF after the outbreak of the epidemic than the model with 60-day lookback.

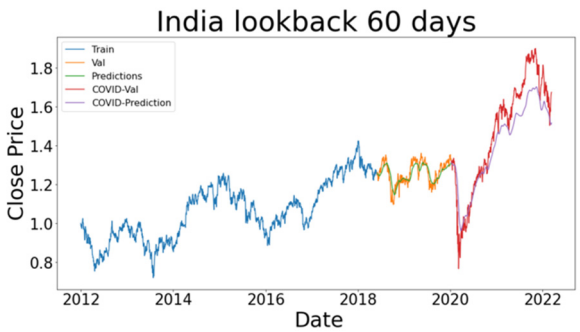

Figure 4 Prediction of India using 60-day lookback

Malaysia is generally similar to India, and the ETF price change trend obtained by the model is roughly the same as the real situation. The difference is that among the 3 levels of lookback, the 15-day lookback had the best performance (total RMSE=3.439, figure 4). However, even the forecast curve seen in 15 days lookback still had a significant gap from the real ETF price change curve.

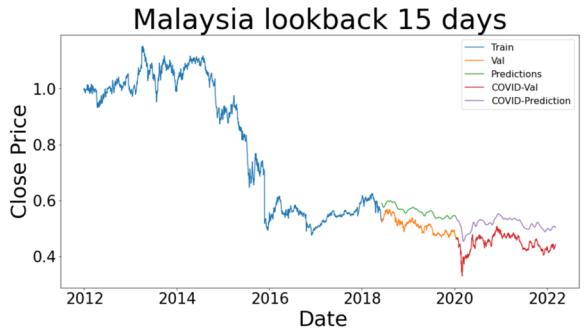

Figure 5 Prediction of Malaysia using 15-day lookback

In addition, the results from the model in the Philippines and Thailand are also very similar. Before the outbreak of the Covid-19, the 60-day lookback and 30-day lookback gave more accurate predictions for the Philippine's ETF (RMSE equal to 0.976 and 1.055 respectively), while the forecast values obtained by the 15-day lookback were slightly lower; after the outbreak of the Covid-19, the forecast curves given by the 15-day lookback was closer to the true values (RMSE=1.477) than the forecast curves obtained by the 60-day lookback and 30-day lookback. Thailand's 30-day lookback is more accurate in predicting changes in ETF after the pandemic (RMSE=3.503), while a 60-day lookback performs better at predicting pre-pandemic ETF changes (RMSE=2.117).

The model with 60 days as a lookback parameter (Figure 5) predicts the trend change of Indonesian ETF prices better than models with 15 days and 30 days as lookback parameters, but the forecast curve after the epidemic is still somewhat different from the real curve.

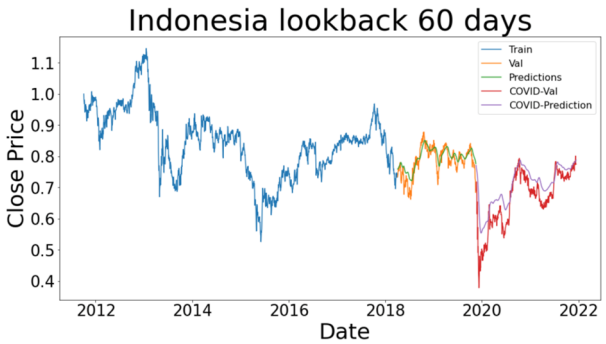

Figure 6 Prediction of Indonesian using 60-day lookback

Overall, as the number of days to lookback becomes larger, the model tends to get a more accurate prediction. However, the actual parameters still need to be determined according to the specific historical stock price movements in different regions. Meanwhile, for the decrease of ETF prices in various regions to varying degrees after the epidemic, the model can predict this trend but still have a certain gap in value from the real situation.

4. DISCUSSION

From the results we got, it is clear that in most cases, using 60-day lookback could get the best results but for some stock markets like the India stock markets. And for the Philippines stock markets, using 15-day lookback could get the best results and for the Thailand stock markets, using 30-day lookback could provide the best result to us. This phenonmen may be caused by the stock market cycle which meant that the price of the stock usually shows periodicity. And when we used quantitative analysis to predict the stock price, we used the historical data to predict the future data.

But it is not that using more days' data to lookback, the better results we could get. There is a phenomenon from our results that for some markets, 60-day lookback could provide the best results while for some markets, the 15-day or 30-day lookback could provide the best results. This is also because of the stock market cycle, there are different stock market cycles in different stock markets. So, for different stock markets, the best number of lookback days is also different. And an example in our results could also prove this for the Philippine stock market. For this market, before the breakout of the outbreak of the epidemic, the 60-day lookback could provide the best result while after the outbreak of the epidemic, the 15-day lookback could provide the best result. It is just because the outbreak of the epidemic could make a great impact on the economy of the country and then the stock market cycle would also be changed. So, when we use the LSTM to predict ETF, the best number of lookback days we use would be influenced by the life cycle of that stock market.

In our research, we also divided the result into two parts. The first is before the outbreak of the epidemic and another is after the outbreak of the epidemic. And it is obvious that in most situations after the outbreak of the epidemic, the results we got become worse than those before the outbreak of the epidemic. The reason for this phenomenon is that the macroeconomy has a great influence on the stock market. Because once the economy in a country has a recession tendency, the investor would lose their heart in this country's stock market and the economic recession may even affect the investors that make them don't have enough cash flow. Since the appearance of the COVID-19, there are many irregular outbreaks of the epidemic in almost all countries of the world and each time it will hit the economy and stock of the countries. It is just because of the uncertainty of the outbreaks of the epidemic, the uncertainty of the stock market also increases and the ETF becomes more difficult to predict. So, after the outbreak of the epidemic in each country, the result becomes worse. We could use the Singapore stock market as an example to prove this. Singapore stock market is a very mature stock market and Singapore Exchange Limited is the fourth largest stock exchange in the world. So, the stability of the Singapore

stock market is very high. The epidemic in Singapore has two heavy outbreaks which were in April 2020 and September 2020 respectively. And it is obvious that in our result of Singapore 60 days lookback, the results in April 2020 and September 2020 are the worst. Overall, the outbreak of the epidemic makes a great difference in the effects of the LSTM.

5. CONCLUSIONS

This paper adopts LSTM to predict the value of ETF of Singapore, India, Malaysia, Philippines, Thailand, and Indonesia respectively, which selects 15-day, 30-day, and 60-day lookback as the parameter to build the model. And for different stock markets, because of the different stock market cycles, the best performance points are in the different number of days lookback. In addition, this study also compares the performance of this model before the outbreak of the COVID-19 epidemic and after the outbreak of the COVID-19 epidemic. The research also finds that the epidemic would also impact the performance of the prediction. As it can't get rid of the COVID-19 epidemic in a short time, maybe the future study needs to find some good models that could help reduce the effect of the epidemic to the prediction in the future.

REFERENCES

[1] S. Borovkova, I. Tsiamas, An ensemble of LSTM neural networks for high-frequency stock market classification, Journal of Forecasting, vol. 38, no. 6, 2019, pp. 600-619. DOI: 10.1002/for.2585

[2] W. Lu, J. Li, Y. Li, A. Sun, J. Wang, A CNN-LSTM-based model to forecast stock prices, complexity, vol. 2020, 2020, DOI: 10.1155/2020/6622927

[3] Sen J, Dutta A, Mehtab S, Profitability Analysis in Stock Investment Using an LSTM-Based Deep Learning Model//2021 2nd International Conference for Emerging Technology (INCET), 2021, pp. 1-9. IEEE

[4] A. Moghar, M. Hamiche, Stock market prediction using LSTM recurrent neural network, Procedia Computer Science, vol. 170, 2020, pp. 1168-1173. DOI: 10.1016/j.procs.2020.03.049

[5] M. A. I. Sunny, M. M. S. Maswood, A. G. Alharbi, Deep learning-based stock price prediction using LSTM and bi-directional LSTM model, 2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES), 2020, pp. 87-92. DOI: 10.1109/NILES50944.2020.9257950

[6] K. Pawar, R.S. Jalem, V. Tiwari, Stock Market Price Prediction Using LSTM RNN, Emerging Trends in Expert Applications and Security, vol. 841, 2019, pp. 493-503. DOI: 10.1007/978-981-13-2285-3_58

[7] O. Wyman, Accelerating Financial Inclusion in South-East Asia with Digital Finance, Asian Development Bank, 2017, from https://think-asia.org/handle/11540/7500

[8] P. Narayan, R. Smyth, M. Nandha, Stock Market Linkages in South–East Asia, East Asian Economic Association 2002, vol. 16, no. 4, 2002, pp. 353-377, DOI:10.1007/978-3-540-30494-4_16

[9] S. Wahyudi, H. Hersugondo, R. D. Lasksana, R. Rudy, Macroeconomic Fundamental and Stock Price Index in Southeast Asia Countries: A Comparative Study, International Journal of Economics and Financial Issues, vol. 7, no. 2, 2017, pp. 182-187, from https://dergipark.org.tr/en/download/article-file/365654

[10] S. Mehtab, J. Sen, A. Dutta, Stock price prediction using machine learning and LSTM-based deep learning models, Symposium on Machine Learning and Metaheuristics Algorithms, and Applications, 2020, pp. 88-106. DOI: 10.1007/978-981-16-0419-5_8

[11] C. Arusha, W. Guneratne, The efficiency of emerging stock markets: Empirical evidence from the South Asian region, 2007, pp.171-183. from https://www.jstor.org/stable/40376165

[12] R. Kumar, R S. Dhankar, Fiancial instability, integration and volatility of emerging South Asian stock markets, South Asian Journal of Business Studies, 2017.

[13] J. Schmidhuber, S. Hochreiter, Long short term memory, Neural Comput, vol. 9, no. 8, 1997, pp. 1735-1780.