Multi-Time Scaled Neural Network Predicting the Moving Average Line of Financial Time Series

(Abstract) This paper proposes a neural network predicting the mid-long-term moving average line of financial time series based on the multi-time scaled feature map of the moving average line from the past. Specifically, the proposed model having a fully connected neural network predicts the moving average line of the next 50 days for the S&P 500 based on the multi-time scaled feature map of the moving average line from the past 121 days. In order to extract an effective representation of past market patterns to predict the moving average line of financial time series, the moving average line from the past is reshaped into a multi-time scaled matrix by using the raster scanning order and encoded into a multi-time scaled feature map by the proposed model having two-dimensional convolutions and two-dimensional max pooling. When evaluating the practical efficacy of the proposed model, backtesting predictions of a unique 50-day time frame in the moving average line of the S&P 500 were considered to be accurate if they had a mean squared error lower than the training mean squared error of the proposed model. Subsequently, a set of data points in the moving average line of the S&P 500 were counted as predicted data points if they were included in at least one of the accurate backtesting predictions made by the proposed model. Empirical results suggests that while the probability of the proposed model outputting an accurate backtesting prediction was observed to be 0.4405, the accurate predictions were able to predict 98.90% of the data points in the moving average line of the S&P 500.

1. INTRODUCTION

Investing money in a variety of securities, such as stocks, funds, bonds, commodities, futures, and cryptocurrency, is one way to gain profit. However, making the best trading decision is challenging due to the volatile and non-linear nature of financial markets. Nevertheless, computational models have been successful in making statistical decisions to be profitable. For instance, a diffusion index [1] and a support vector machine [2] are able to predict the return of a stock by using a set of mathematical formulas that derive important features in market patterns, such as market cycles [1] and the periodic relationships and change in price data [2]. However, unlike market forecasting models using a set of mathematical formulas, recent studies [3, 4] followed by the rise of deep learning are starting to dump financial time series into black boxes without reducing noise and extracting meaningful features. Since the purpose of utilizing deep learning models is to automatically learn the features from the data, it is reasonable that the results were observed to be satisfactory. However, deep learning models would be limited to short-term predictions because a sequence of data points that ineffectively represent the sequential and periodic features of past market patterns wouldn't provide enough useful information to predict the mid-long-term future. In order words, the lack of a proper method of feature extraction is preventing deep learning-based market forecasting models from making more complex and longer predictions. Therefore, a deep learning model predicting the non-linear trend of financial time series in the mid-long-term future must have an effective method of feature extraction that encodes the sequential and periodic features of past market patterns.

In this paper, a deep learning model having a method of two-dimensional feature extraction is proposed for predicting the non-linear trend of financial time series in the mid-long-term future. Specifically, the proposed model having two-dimensional convolutions and two-dimensional max pooling reshapes the moving average line of financial time series from the past into a multi-time scaled matrix by using the raster scanning order and encodes the sequential and periodic features in the multi-time scaled matrix of the past into a multi-time scaled feature map. Subsequently, the multi-time scaled feature map of the past is given to the neural network of the proposed model to predict the moving average line of financial time series in the mid-long-term future. Unlike other market forecasting models that lacks a proper method of feature extraction, the proposed model having two-dimensional convolutions and two-dimensional max pooling is able to encode the sequential and periodic features in the moving average line from the past into an effective representation of past market patterns. Consequently, having an effective representation of past market patterns enables the neural network of the proposed model to efficiently decode the multi-time scaled features of the past and therefore generate more complex and longer predictions like the moving average line of financial time series in the mid-long-term future.

Being able to predict the moving average line of a financial time series in the mid-long-term future is expected have an investment benefit other market forecasting models do not have. Specifically, most market forecasting models are designed to predict the one-day directional movement or the buy-and-hold return of a security. Unlike such market forecasting models, the proposed model predicting the moving average line of financial time series would be able to provide a more complex anticipation of the ups and downs of the future market that would help decide which securities to invest in and how to optimize an investment portfolio.

The remainder of this paper is organized as follows. Section 2 reviews the method of other deep learning-based market forecasting models and their weakness the proposed model attempted to improve on. Section 3 explains the method of the proposed model. Section 4 presents the empirical results that were obtained by backtesting the proposed model. Section 5

본 논문에 포함된 정보의 전부 또는 일부를 무단으로 제3자에게 공개, 배포, 복사 또는 사용하는 것을 엄격히 금지합니다.

discusses the obtained data. Section 6 concludes this paper with a summary of the method, results, applications, and risks of the proposed model. Finally, Section 7 ends this paper by briefly suggesting potential adjustments that can be made to improve the practical efficacy and application of the proposed model.

2. RELATED WORK

Before explaining the method of the proposed model, this section will briefly introduce other deep learning-based market forecasting models with a weakness the proposed model attempted to improve on.

Fischer and Krauss [3] proposed that a long-short-term memory (LSTM) neural network is able to predict the one-day return of securities included in the S&P 500. Specifically, they used a sequence of one-day returns of a security included in the S&P 500 as an input fed into a LSTM network predicting the next day's return of that security. Although their model was reported to have a Sharpe Ratio of 5.8 and daily returns of 0.46% before fees, a major issue with their model is that the sequence of one-day returns of a security is immediately fed into the LSTM network without extracting meaningful features. Not having a method of extracting features from the sequence of one-day returns is an issue because their model would be given an input full of noise. Moreover, their model may miss important information from the past because LSTM networks are designed to sequentially observe the input and therefore focus on the sequential features from the past when there are also periodic relationships that are important for analyzing past market patterns and their prolonging effect. Consequently, not having a proper method of feature extraction and the tendency for LSTM networks to reduce the significance of past data points over recent data points may not provide Fischer and Krauss' model with enough useful information of past market patterns to generate predictions that are more complex and longer than predicting short-term returns.

Guo et al. [4] proposed a model that improves Fischer and Krauss' model but still exhibits a similar issue with feature extraction. Like Fischer and Krauss, the model proposed by Guo et al. predicts the one-day return of bitcoin prices by using a LSTM network. However, Guo et al. used one-dimensional convolutions with varying window sizes to extract the sequential features in past price data that are fed into their LSTM network predicting the next day's price. Although their LSTM model does have a method of feature extraction, concatenating a combination of sequential features in past price data does not take account of the fact that past market patterns are a complex combination of sequential and periodic features. Therefore, the model proposed by Guo et al. may not be able to have an effective representation of past market patterns that are needed to generate predictions that are more complex and longer than predicting short-term returns.

In short, previous deep learning-based market forecasting models exhibited a weakness in extracting an effective representation of the sequential and periodic features in past market patterns that should be used to generate more complex and longer predictions than predicting short-term returns. Such a weakness in feature extraction that limited previous models to predictions with low complexity became an inspiration to implement and propose a neural network that can predict the moving average line of financial time series in the mid-long-term future based on a multi-time scaled feature map of the past extracted with two-dimensional convolutions and two-dimensional max pooling.

3. METHOD

The proposed model was designed to predict the 10-day moving average line of the next 50 days for the S&P 500 (SPY) based on the 50-day moving average line from the past 121 days. The overall architecture of the proposed model is outlined in Figure 1.

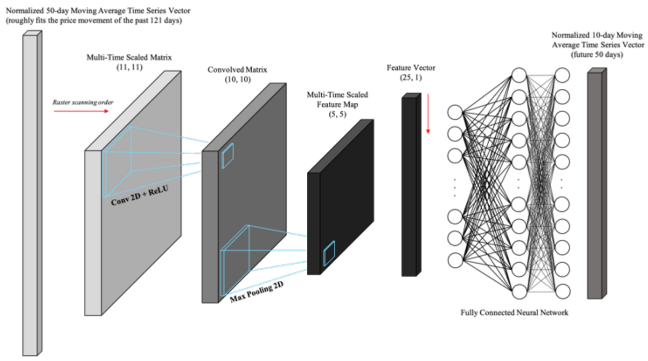

Fig. 1. The proposed model architecture. The proposed model having two-dimensional convolutions and two-dimensional max pooling encodes the moving average line from the past 121 days into a multi-time scaled feature map. The neural network of proposed model then uses the multi-time scaled feature map of the past to predict the moving average line of the next 50 days.

3.1 DATA SAMPLING

To sample the dataset used to train and test the proposed model, a time frame observing 180 data points, denoted as t, was slid through the historical price data of SPY. Within t, one input consisting of 121 data points and its corresponding output consisting of 50 data points are sampled. Specifically, the input sampled in t is the min-max normalized 50-day moving average line that roughly fits the first 121 data points in t, whereas the output sampled in t is the min-max normalized 10-day moving average line that roughly fits the price movement of 50 data points in t succeeding the 121st data point in t. Equations 1 and 2 was used to sample and normalize the n-day moving average line of a given time frame in historical price data of SPY.

$$m_i = \sum_{k=i}^{i+n-1} t_k \quad (1)$$

$$m_i^{norm} = \frac{m_i - m_{min}}{m_{max} - m_{min}} \quad (2)$$

Note that a 50-day moving average was used to indicate the long-term trend of the past with reduced noise, whereas a 10-

본 논문에 포함된 정보의 전부 또는 일부를 무단으로 제3자에게 공개, 배포, 복사 또는 사용하는 것을 엄격히 금지합니다.

day moving average was used to sample some noise indicating minor corrections and major drop downs the proposed model intends to predict. Implementing this method of sampling acquired 5,104 samples from SPY between January 1, 2000 and January 1, 2021. An example of a sampled input and output are shown in Figure 2.

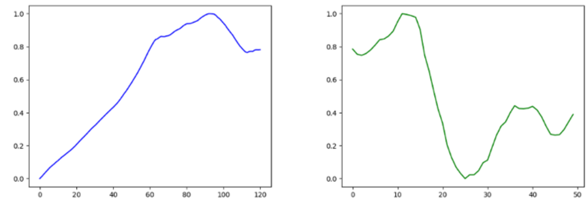

Fig. 2. An example of the 50-day moving average line from the past 121 days (left) and the 10-day moving average line of the next 50 days (right). The graph in the left (blue line) is an example of a long-term trend from the past that is used by the proposed model to predict future market trends like the graph in the right (green line).

3.2 MULTI -TIME SCALED FEATURE EXTRACTION

The moving average line from the past must be condensed into an effective feature map in order to enable the prediction of the moving average line in the mid-long-term future. Thus, to effectively extract the features of past market patterns, the raster scanning order was used to reshape the moving average line from the past 121 days into a 11x11 matrix. Using the raster scanning order to reshape the moving average line from the past 121 days into a 11x11 matrix creates a multi-time scaled matrix of the past having columns of sequential data points and rows of periodic data points. Subsequently, 2x2 convolutions with a stride of 1 and ReLU was used to encode the sequential features between data points at adjacent columns and the periodic features between data points at adjacent rows in the multi-time scaled matrix of the past. Once the multi-time scaled matrix of the past is convolved, 2x2 max pooling was used to condense and smooth out the extracted features of past market patterns. As a result, the moving average line from the past 121 days is encoded into multi-time scaled feature map as shown in Figure 3. After the proposed model encodes the multi-time scaled feature map of the past, the raster scanning order was used once more to flatten the multi-time scaled feature map into a feature vector that fits the neural network of the proposed model.

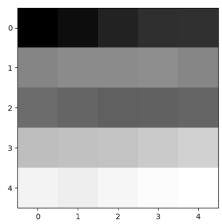

Fig. 3. 5x5 multi-time scaled feature map of a moving average line from the past 121 days. The moving average line from the past is encoded into a meaningful noise representing the sequential and periodic features of past market patterns. Note that darker cells have a value closer to 0.0 whereas brighter cells have a value closer to 1.0 or higher.

Note that having one layer of convolutions and max pooling condenses the moving average line from the past 121 days into a feature map with 25 data points as shown in Figure 3. Therefore, only one layer of convolutions and max pooling was used because having too many layers of convolutions and max pooling may result in losing important information in the moving average line from the past by oversimplifying the extracted features.

3.3 NEURAL NETWORK

Given the 5x5 multi-time scaled feature map of the moving average line of SPY from the past 121 days, the moving average line of the next 50 days for SPY is predicted by the proposed model having a fully connected neural network. Since most of the feature extraction is completed by encoding the multi-time scaled feature map of the past, the neural network of the proposed model focuses on decoding the extracted features to predict the moving average line of SPY in the mid-long-term future.

The architecture of the neural network of the proposed model is explained as follows. The neural network of the proposed model has one hidden layer with 50 outputting nodes reading 25 input features from the 5x5 multi-time scaled feature map of the past that is flattened into a feature vector by using the raster scanning order. The output layer with 50 outputting nodes that follows the hidden layer generates a prediction of the moving average line of the next 50 days for SPY. All outputting nodes in the neural network uses ReLU as the non-linear activation function as written in equation 3. Note that x_j either represents the input or the previous layer's activations.

$$o_i = \max\left(0, \sum_{j=0}^{n} x_j w_{ij}\right) \quad (3)$$

The mean squared error (MSE) between the supervised output and the prediction made by the proposed model was used as the cost function of the neural network as written in equation 4. To optimize the neural network, stochastic gradient descent with a step decay learning rate was implemented as written in equation 5. Note that the learning rate denoted as α decays by a factor r for every N training epoch.

$$E = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (4)$$

$$w_{ij}^{t+1} = w_{ij}^{t} - \frac{\partial E}{\partial w_{ij}^{t}} \cdot \alpha(1-r)^{\frac{[t]}{N}} \quad (5)$$

3.4 IMPLEMENTATION

The first 85% of the sampled dataset (4339 samples) was randomly shuffled and used to train the proposed model. The proposed model was trained for 10,000 epochs with an initial learning rate of 0.01 that decayed by 1% per 1,000 epochs.

본 논문에 포함된 정보의 전부 또는 일부를 무단으로 제3자에게 공개, 배포, 복사 또는 사용하는 것을 엄격히 금지합니다.

The last 15% of the sampled dataset (765 chronological samples) was used for backtesting. Note that for each backtesting sample, the proposed model predicted a unique 50-day time frame in the moving average line of SPY having 815 data points.

The proposed model was implemented in C++ and Python without using any external frameworks. The code and full backtesting results of the proposed model are open-sourced in github.com/junyoung-sim/sltm.

4. RESULTS

Table 1 shows the training and backtesting MSE of the proposed model.

Table 1. Training and Backtesting Mean Squared Error of the Proposed Model

Type	Mean Squared Error (MSE)
Training	0.0543546
Backtesting	0.0940630

Since the backtesting MSE was observed to be higher than the training MSE, a backtesting prediction of a unique 50-day time frame in the moving average line of SPY was considered as an accurate prediction if it had a MSE lower than the training MSE of the proposed model. Figures 4 through 7 shows several accurate backtesting predictions made by the proposed model in chronological order.

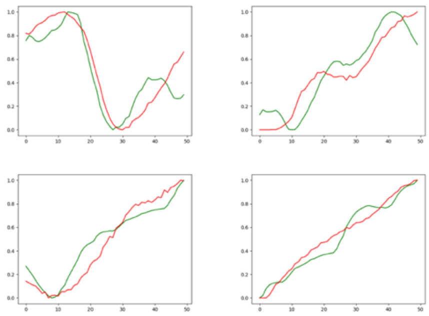

Fig. 4. Example of accurate backtesting predictions #1. The proposed model made accurate predictions of a unique 50-day time frame in the moving average line of SPY between January of 2018 and June of 2018.

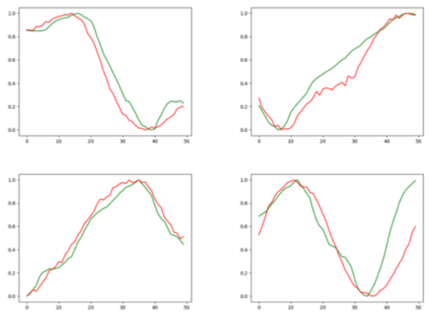

Fig. 5. Example of accurate backtesting predictions #2. The proposed model made accurate predictions of a unique 50-day time frame in the moving average line of SPY between September of 2018 and July of 2019.

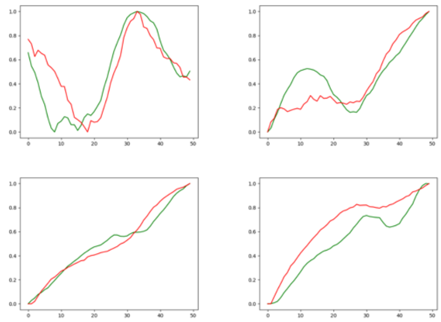

Fig. 6. Example of accurate backtesting predictions #3. The proposed model made accurate predictions of a unique 50-day time frame in the moving average line of SPY between August of 2019 and January of 2020.

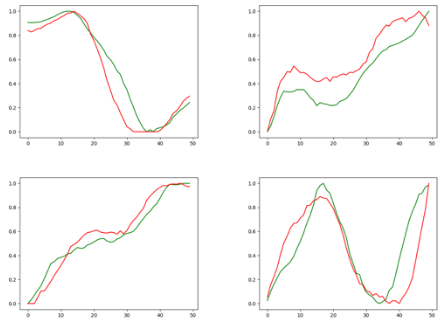

Fig. 7. Example of accurate backtesting predictions #4. The proposed model made accurate predictions of a unique 50-day time frame in the moving average line of SPY between February of 2020 and December of 2020.

Based on the distribution of backtesting predictions shown in Figure 8, it was observed that 337 predictions out of 765 chronological backtesting samples were accurate predictions like the samples shown in Figures 4 through 7 that have a MSE lower than the training MSE of the proposed model. Therefore, the probability of the proposed model outputting an accurate prediction was observed to be 0.4405.

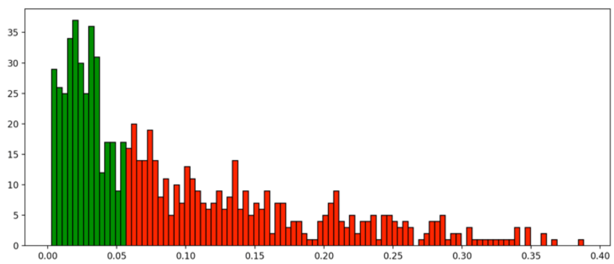

Fig. 8. Distribution of backtesting predictions. The green bins in the histogram indicate the distribution of accurate backtesting predictions having a MSE lower than the training MSE of the proposed model that was observed to be 0.0543546.

More data was collected from the accurate predictions to evaluate the practical efficacy of the proposed model. The accurate predictions would be effective only if they are generated by the proposed model in a regular basis such that a continuous insight into future trends of financial time series is provided without leaving a certain time frame from being

 본 논문에 포함된 정보의 전부 또는 일부를 무단으로 제3자에게 공개, 배포, 복사 또는 사용하는 것을 엄격히 금지합니다.

unpredicted. Therefore, the practical efficacy of the proposed model was evaluated by counting the i th data point in the moving average line of SPY as a predicted data point if it was included in at least one of the 337 accurate backtesting predictions. Under this method of evaluating the practical efficacy of the proposed model, a set of 806 data points out of 815 data points (98.90%) in the moving average line of SPY were observed to be predictable by the proposed model. A visual representation of the frequency of outputting accurate predictions and the set of predictable data points are shown in Figure 9 and 10.

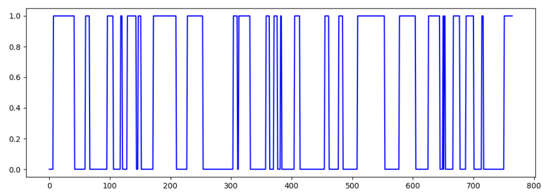

Fig. 9. Frequency of outputting accurate backtesting predictions. To visualize how regularly the proposed model made accurate predictions, the n th backtesting prediction of a unique 50-day time frame in the moving average line of SPY was plotted as 1.0 if it had a MSE lower than the training MSE and plotted as 0.0 otherwise.

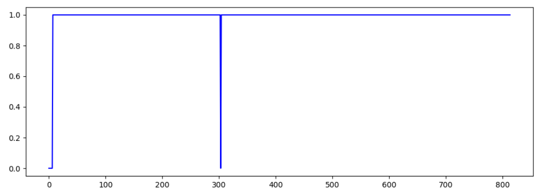

Fig. 10. Predicted data points. The i th data point in the moving average line of SPY was plotted as 1.0 if predicted by at least one of the 337 backtesting predictions having a MSE lower than the training MSE and plotted as 0.0 otherwise.

5. DISCUSSION

An accurate prediction of a unique 50-day time frame in the moving average line of SPY must be able to approximate the general direction and magnitude of change in the trend. Therefore, it was plausible to consider backtesting predictions having a MSE lower than the training MSE to be accurate predictions. Figures 4 through 7 indeed shows that predictions having MSE lower than the training MSE accurately approximated the direction and magnitude of change in the moving average line of SPY at various market seasons, such as gaining seasons, minor corrections, and major drop downs. Nevertheless, one may assume that the proposed model would be ineffective since the probability of outputting an accurate prediction of a unique 50-day time frame in the moving average line of SPY was observed to be lower than 0.5. However, Figure 9 shows that the proposed model consistently made accurate predictions of a unique 50-day time frame in the moving average of SPY except the first few predictions and the predictions between the 250th prediction and the 300th prediction labeled as 0.0. As a result, it was observed that consistently generating accurate predictions enabled the proposed model to accurately approximate the direction and magnitude of change in 98.90% of the data points in the moving average line of SPY as shown in Figure 10. The observed results are indeed logical because the frequency of outputting accurate predictions would be more significant than the number of accurate predictions since the proposed model consistently made at least one accurate prediction per 50 days with each prediction forecasting a unique 50-day time frame in the moving average line of SPY.

6. CONCLUSION

The lack of a proper method of feature extraction have limited previous deep learning-based market forecasting models to predicting short-term return of a security. Thus, this paper proposed a model having a neural network predicting the moving average line of financial time series in the mid-long-term future based on the multi-time scaled feature map of the moving average line from the past encoded with two-dimensional convolutions and two-dimensional max pooling. Empirical results obtained by backtesting the proposed model suggests that the probability of the proposed model outputting an accurate prediction of a unique 50-day time frame in the moving average line of SPY was lower than 0.5. However, the frequency of outputting accurate predictions was observed to be more significant than the number of accurate predictions as additional data suggests that the proposed model was able to approximate the direction and magnitude of change in 98.90% of the data points in the moving average line of SPY by consistently generating at least one accurate prediction per 50 days. The observed results as a cohesive whole ultimately suggest that a continuous insight into non-linear mid-long-term trends of financial time series can be generated by a neural network that is given a multi-time scaled feature map of the past encoded with two-dimensional convolutions and two-dimensional max pooling. In terms of its practical application, the proposed model showed the capability to benefit mid-long-term portfolio optimization and help decide whether or not to invest in a security by predicting the non-linear trend of various market seasons that are outside the scope of other market forecasting models predicting the buy-and-hold return of a security. Using the proposed model in practice would still require some consideration since the backtesting results presented in this paper may not apply to future performance if drastic changes in market behavior contradicts with patterns the proposed model learned from the past. Thus, consistent training and testing must be conducted in order to implement the proposed model in actual practice.

7. FUTHER WORK

7.1 Prediction Confidence

The proposed model does not have a method of predicting whether a prediction will be an accurate prediction having a MSE lower than the training MSE. Therefore, a statistical method evaluating the confidence of predictions must be implemented in order to strengthen the practical application of the proposed model.

7.2 MULTIVARIATE TIME SERIES

Since the proposed model is dependent on past data, it may underperform than the backtesting results discussed in this paper if the market changes its behavior due to economic, social, political, and other factors. The proposed model can be improved to prevent this risk by encoding the multi-time scaled feature map of multiple economic indicators as different channels that are used to predict the mid-long-term moving average line of a security.

7.3 HYPERPARAMETERS

The hyperparameters used in this study were acquired through rigorous testing. However, additional testing with different time intervals for the price data, different time series sizes, different convolution and pooling dimensions, and different neural network architectures would be necessary to implement the proposed model in a wider range of market forecasting perspectives.

7.4 WIDER APPLICATIONS

This paper only introduced an implementation of the proposed model on SPY. Testing on other economic indicators and financial assets would be necessary to verify the efficacy of the proposed model in a wider spectrum. Additional work may also be focused on whether the proposed model can be applied in other fields of study such as signal processing and image compression.

REFERENCES

[1] Stock, J.H. & Watson, M. W. Macroeconomic Forecasting Using Diffusion Indexes. *Journal of Business & Economic Statistics* **20**, 147-162 (2002)

[2] Kim, K. Financial time series forecasting using support vector machines. *Neurocomputing* **55**, 307-319 (2003)

[3] Fischer, T. and Krauss C. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research* **270**, 654-669 (2018)

[4] Guo, Q. et al. MRC-LSTM: A Hybrid Approach of Multi-scale Residual CNN and LSTM to Predict Bitcoin Price. *International Joint Conference on Neural Networks (IJCNN)*, 1-8 (2021)

 본 논문에 포함된 정보의 전부 또는 일부를 무단으로 제3자에게 공개, 배포, 복사 또는 사용하는 것을 엄격히 금지합니다.