# PREDICTION AVERAGE STOCK PRICE MARKET USING LSTM



Rezzy Eko Caraka
Chaoyang University of Technology

# PREDICTION AVERAGE STOCK PRICE MARKET USING LSTM

**CHUN YUAN LAI, RUNG-CHING CHEN*, REZZY EKO CARAKA[1]**

College of Informatics, Chaoyang University of Technology, Taichung City, 41349, Taiwan (R.O.C.)
E-MAIL: john22472417@gmail.com, *crching@cyut.edu.tw, rezzyekocaraka@gmail.com

**Abstract:**

**Predicting stock price has been a challenging project for many researchers, investors, and analysts. Most of them are interested in knowing the stock price trend in the future. To get a precise and winning model is the wish of them. Recently, Neural Network is a prevalent way to be the prediction. However, there are many ways and different predicting projects such as Convolutional Neural Networks (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU). In this paper, we propose a novel idea that average previous five days stock market information (open, high, low, volume, close) as a new value then use this value to predict, and use the predicted value as the average of the stock price information for the next five days. Moreover, we are then based on the Technical Analysis Indicators to consider whether to buy stocks or continue to hold stocks or sell stocks. We use Foxconn company data and collect from Taiwan Stock Exchange by using the Neural Network Long Short-Term Memory (LSTM).**

**Keywords:**

**LSTM, Stock price prediction, Technical Analysis Indicator.**

## 1. Introduction

In recent years, Neural Network is popular ways to be the prediction. However, there are many ways such as Convolutional Neural Networks (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU). The idea of RNN is straightforward, which is to help us process information in stages. When referring to a standard Neural Network all inputs and outputs are not dependent on each other. Only if the researcher uses this concept, what happens is that the tasks on the neural network are very numerous and stacked up. Whereas if researchers want to build a forecasting system, we need a prediction from the best input so that the model can learn well from historical data. Most researchers use a significant lag from the Box Jenkins time series model to be used as input. RNN performs the same task on each element in a sequence, then processes the output that refers to previous computing.

Also, different predicting projects, for example, object detection, face detection, Brain wave signal identification. One of them and has always been a favorite research project is stock price prediction. Moreover, its development can be applied to other time-series forecasting problems, such as crowd flow forecasting[2], clinical medical [3], weather [4], rainfall [5], wind speed forecasting [6], electrical load forecasting [7] [8], human behaviour forecasting [9] , extreme event forecasting [10], cargo forecasting[11] and traffic flow forecasting [12] .

People want to invest stocks because they hope to increase their investment benefit. Many factors affect the stock price movement such as inflation rates, economic environment, political issue, and so on [13]. However, the company's poor management will lead to a decline in stock prices, so investors hope that there can be high-accuracy and no overfitting models to predict stock prices and understand the stock prices trend of future. Before the use of neural networks to predict stock prices, most people use Technical Analysis Indicators to evaluate the current state of stocks, suitable for buying, selling and continuing to hold. In this paper, we use Stochastics oscillator (KD), Moving Average Convergence / Divergence (MACD), Relative Strength Index (RSI) and On Balance Volume (OBV) this four technical analysis indicators to evaluate based on our predict stock prices state suitable for buying, selling, and continuing to hold. The purpose of this paper is to explore the prediction for the value to be average in future five days that use average previous five days of stock market information. Foxconn is select from the Taiwan Stock Market. A Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) applied in this research. The paper organized as follows. In section II, presents the neural network RNN and LSTM structure and Technical Analysis Indicators for KD, MACD, RSI, and OBV. Section III data sources and the experiment result shown in section IV, regarded as the conclusion.

## 2. Methodology

To facilitate understanding of the method in this paper, we briefly introduce the structure of RNN and LSTM, besides, discuss why LSTM excels in sequence analysis. Although RNNs have been proven successful in sequence prediction tasks, it can still be challenging to learn the long-term dependence, mainly due to the explosion /vanishing gradient problem that results from the gradient propagation of the recurrent network over many layers.

## 2.1. Recurrent Neural Network (RNN)

Jeff Elman [13] proposed a Recurrent Neural Network (RNN). Mainly, RNN solves the problem of processing of sequence data, such as text, voice, and video. There is a sequential relationship between the samples of this type of data, and each sample is associated with its previous sample. For example, in the text, a word related to the word in front of it. In meteorological data, the temperature of a day associated with the temperature of the previous few days. A set of observations defined as a sequence from which multiple sequences can be observed. In Fig. 1 the output of the hidden layer is stored in the memory. Memory can be considered as another input. The main reason for the difficulty of RNN training is in the propagation of the hidden layer parameter ω. Since the error propagation is on the unrolled RNN, ω multiplies multiple times during both the forward propagation process and the backpropagation process. (1) Gradient Vanishing problem is when the gradient is small, increasing many exponential drops/, there is almost no effect on the output. (2) Gradient explosion problem: Conversely, if the gradient is large, multiplying multiple exponential increases leads to a gradient explosion. Of course, this problem exists in any deep neural network, but it is especially apparent due to the recursive structure of RNN.

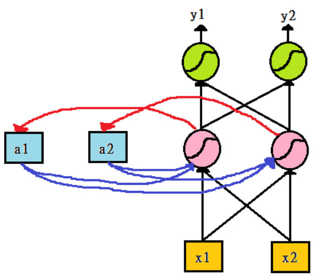


**FIGURE 1.** The Recurrent Neural Network Structure.

## 2.2. Long Short-Term Memory (LSTM)

Hochreiter and Schmidhuber invented Long Short-Term Memory (LSTM) networks in 1997 [14]. Which is uniquely different from traditional RNN, is different from the construction of neurons. There is no difference between each neuron in traditional RNN and general neural network. However, in LSTM, each neuron is a "memory cell" that connects previous information to the current mission. There are three gates in LSTM, input gate, forget gate, and output gate. Activation function f is usually a sigmoid function between 0 and 1.

Fig. 2 explain when the output of a certain part of the neural network (the outside world) wants to be written into the memory cell, it must first be input to the memory cell through the input gate. First, If $zi$ passes the activation function $(zi)$ is 0, time with input $(z)$, $(z)$ $(zi)$ will get 0. On the other hand, If $zi$ passes the activation function $(zi)$ is 1, time with input $(z)$, $(z)$ $(zi)$ will get value. Second, If $zf$ passes the activation function $(zf)$ is 0, time with memory cell c, c $(zf)$ will get 0 and memory cell c will forget the value last time. Namely, if $zf$ passes the activation function $(zf)$ is 1, time with memory cell c, c $(zf)$ will get value and memory cell will remember the value last time. Then, c′ equal $g(z)f(zi)$+$cf(zf)$. Third, if $zo$ passes the activation function $f(zo)$ is 0, time with hidden layer h$(c')$, h $(c')f(zo)$ will get 0 and will be no output. Namely, if $zo$ passes the activation function $f(zo)$ is 1, time with hidden layer h $(c')$, h $(c')f(zo)$ will get value and will have output.

Our proposed method, the data fed to the neural network and trained for prediction assigning random biases and weights. Our LSTM model is composed of a sequential input layer followed by 2 LSTM layers and dense layer with Relu activation and then finally a dense output layer with linear activation function.

$$g(z)f(z_i) \tag{1}$$

$$cf(z_f) \tag{2}$$

$$c' = g(z)f(z_i) + cf(z_f) \tag{3}$$

$$a = h(c')f\ (z_o) \tag{4}$$

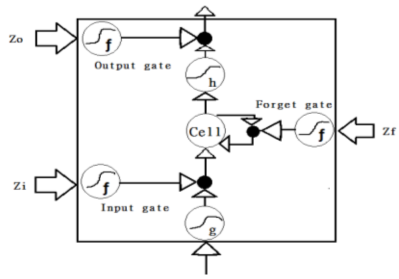


**FIGURE 2.** The Long Short Term Memory structure.

## 2.3. Stochastics oscillator (KD)

The Stochastic Oscillator (KD), formerly known as *%K&%D*. When *%K* the number of periods for slow and *%D* the number of periods for the moving average. Moreover, Users can set the look-back period. KD is also a momentum analysis method in technical analysis, using the concept of overbought and oversold, promoted by George C. Lane in the 1950s. The indicator predicts the timing of the reversal of the price trend by comparing the range of fluctuations in the closing price and price. The term "random" refers to the position of a price over a period of time relative to its range of fluctuations. First, calculate the Raw Stochastic Value (RSV) :

$$RSV = 100\% \times \frac{C_{n-}L_n}{H_{n-}L_n} \quad (5)$$

With n: is the passing transaction period. *Cn*: is the close price of the nth day. *Hn* and *Ln*: the highest and lowest prices in the past *n* days.

Second, find the *K* and *D* values of the day:

$$\mathrm{K}_n = \alpha \times RSV_n + (1-\alpha) \times K_{n-1} \quad (6)$$

$$\mathrm{D}_n = \alpha \times K_n + (1-\alpha) \times D_{n-1} \quad (7)$$

General setting α= 1 / 3.

The *K* value is a "fast average," and the response is sensitive. The *D* value is "slow average" and the response is less sensitive. If the *K* value is > *D* value, it means that it is in the uptrend; otherwise, it is in the downtrend. *K* value and *D* value, the value is between 0% and 100%, 50% is the long-short balance position, 80% is the "Overbought Zone" (overbought zone), the bulls are strong; 20% or less is "oversold" "Oversold Zone," the bears are strong.

## 2.4. Moving Average Convergence / Divergence (MACD)

The Moving Average Convergence / Divergence (MACD) is a common technical analysis tool in stock trading. It was proposed by Gerald Appel in the 1970s to study the strength, direction, energy, and the trend cycle, to grasp the timing of stock buying and selling. The MACD indicator consists of a set of curves and graphs, calculated by the difference between the fast-changing stock price or index and the slowly changing exponential moving average (EMA) at the close. "Fast" refers to EMA in a shorter period, while "Slow" refers to EMA in a more extended period. The most commonly used EMA is 12 and 26 days

$$DIF = EMA_{(close,12)} - EMA_{(close,26)} \quad (8)$$

First, to calculate DIF value use the exponential moving average of the closing price (12/26) to calculate the difference. The signal line (DEM value, also known as MACD value): After calculating the DIF, a "signal line" will be drawn, usually the DIF's 9-day exponential moving average.

$$DEM = EMA_{(DIF,9)} \quad (9)$$

When the DIF or MACD value is greater than 0, it is generally regarded as a long market; otherwise, when the DIF or MACD value is less than 0, it can be regarded as a short market.

## 2.5. Relative Strength Index (RSI)

The Relative Strength Index (RSI) is a technical analysis tool that compares price movements to express price intensity.

The technical analysis method proposed by RSI in June 1978 by American mechanical engineer Welles Wilder JR. was published in the US magazine "Commodities" (now "Futures" magazine) and was included in the same year and introduced in the book New Concepts in Technical Trading Systems. Compared to other analysis tools, RSI is one of the more convenient measurement tools for mass interpreting, so it is prevalent. Let the daily change to *U* and the downward change to *D*. On days when prices rise: *U* = the daily closing price - yesterday's closing price; *D = 0* on a day when prices fall: *U = 0; D* = yesterday's closing price - is the daily closing price.

$$RS = \frac{EMA_{(U,n)}}{EMA_{(D,n)}} \quad (10)$$

$$RSI = (1 - \frac{1}{1+RS}) \quad (11)$$

RS: Relative Strength; RSI: Relative Strength Index; EMA(*U*, *n*): *U* at *n* Average indexes of the day; EMA(*D*, *n*):*D* at *n* Average indexes of the day. According to Wilder's measurement, he pointed out that when the RSI of security rises to 70, it means that the securities have been overbought and investors should consider selling the securities. Conversely, when the securities RSI falls to 30, the representative securities are oversold, and the investor should purchase the securities.

## 2.6. On Balance Volume (OBV)

The On Balance Volume (OBV) was proposed by Joe Granville in the 1960s and is widely used. The four major elements of stock market technical analysis: price, quantity, time and space. The OBV indicator is a technical indicator that uses the "quantity" factor as a breakthrough to discover hot stocks and analyze stock price movement trends. It is to digitize and visualize the popularity of the stock market – the relationship between volume and stock price, and measure the driving force of the stock market by changing the trading volume of the stock market, to judge the trend of the stock price. Regarding the research on volume, the OBV energy tide indicator is one of the most critical analytical indicators. The basis for the establishment of the energy tide theory as follows.

1. The more inconsistent the investor's comments on the stock price, the larger the volume; on the contrary, the volume is small. Therefore, the volume of trading can be used to judge the popularity of the market and the strength of both long and short.

2. Gravity principle. The rising object will fall sooner or later, and the energy required for the object to rise will be more than when it lands. When it comes to the stock market, it can be explained as follows. On the one hand, the stock price will drop sooner or later; the energy required for the stock price to rise is large, so the stock price rise. Notably, at the beginning

of the increase, must have a large volume; It does not have to consume much energy, so the amount does not necessarily enlarge, and there is even a shrinking trend.

3. The principle of inertia - moving is constant, static is static. Only those hot stocks that are investors or the main players will have large fluctuations in volume and stock price in an extensive period, while unpopular stocks will compare the volume and stock volatility over a period of time. If today's closing price is higher than the closing price of the previous trading day, Today's OBV value = OBV value of the last day trading + today's trading volume. On the other hands, if today's closing price is lower than the closing price of the previous day trading, Today's OBV value = OBV value of the last day trading - today's trading volume.

## 3. Experimental Results

Dataset description: We acquired the data from http://www.twse.com.tw/zh/. We have collected the historical stock data of Foxconn, Taiwan Cement Corp., Taiwan Semiconductor Manufacturing Company, Formosa Petrochemical Corporation and Quanta Computer Inc. from the Taiwan stock exchange. We have collected daily dataset and kept a window size of 20 days. The data range is from 01.01.2009 to 31.12.2018. Sequence data: We got 2474 sequences from 01.01.2009 to 31.12.2018, and we changed to average five days to value, so we used 494 data for input. From this dataset, we used 383 samples for training and 43 samples for validation.

To evaluate the stock we use the following way:

1. Stochastics oscillator (KD)

(1) *K* value > 80, more than 3 days, for high-grade passivation, indicating this stock is very strong, the chance of rising again will usually become very high, continue to hold.
(2) *K* value <20, more than 3 days, low passivation, indicating this stock is very weak, the chance to fall again will usually become very high, continue to hold.
(3) The *D* value is between 15 and 85 and continues to be held.
(4) The *K* -line or the *D*-line has not passed or is lower than each other for more than three days, continues to hold.
(5) The *K* line falls below the *D* line, which is the sell signal.
(6) D value >= 85, serious overbought, suitable for selling.
(7) When the *K* line breaks through the *D* line, it is the buy signal.
(8) *D* value <=15, severe oversold, suitable for buying.

2. Moving Average Convergence / Divergence (MACD)

(1) The MACD line or the DIF line has not passed or is lower than each other for more than three days, continues to hold.
(2) The DIF line passes through the MACD line from top to bottom, is a selling signal.
(3) The DIF line passes through the MACD line from bottom to top ,is a buying signal.

3. Relative Strength Index (RSI)

(1) RSI value between 10 and 90, continue to hold
(2) RSI value >=90, severely overbought, suitable for sale.
(3) RSI value <=10, severe oversold, you can buy.

4. On Balance Volume (OBV)

(1) The OBV line changes from rising to falling, indicating that the purchasing power has gradually weakened so that it can be sold at an opportunity.
(2) The OBV line has skyrocketed (continuously rising 3 times), which is a selling signal and sold immediately.
(3) When the OBV line turns from a downtrend to a rising trend, it indicates that the buyer's relative advantage is gradually strengthened. If investors do not buy at this time, the stock price will rise in the future, which will increase the purchase cost.
(4) The OBV line plummeted (continuously fell 3 times) and is a buying signal.

In figure 3, Training Detail: For training the model we choose the feature from open, high, low, close, and volume.

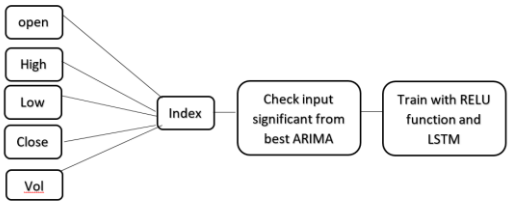


**FIGURE 3.** Flowchart analysis

Based on table 1, it can be seen that the best ARIMA is based on a different stock price index. In general, the advantages of the ARIMA model are that it has a flexible nature (following a data pattern), has a high level of forecasting accuracy and is suitable to be used to predict a number of variables quickly, simply, accurately because it only requires historical data to forecast. Based on the best ARIMA, it will be used as input for the model in LTSM. Moreover, the fractional difference parameter estimate is around -1.0, the data support the deterministic trend model; if the parameter estimate is around 0.0, the data support the unit root with drift model. Model Random Walk with drift is a simple forecasting model used as benchmark.

**TABLE 1.** ARIMA Evaluation

| Stoxx | BEST ARIMA | Parameter | Accuracy | |
|---|---|---|---|---|
| | | | MAPE | RMSE |
| Quanta | ARIMA(2,1,0) | AR1 | 2.34524 | 1.90560 |

| | | | | |
|---|---|---|---|---|
| | | 0.1982 AR2 -0.0893 | | |
| Formosa | ARIMA(0,1,1) | MA 0.1308 | 1.817912 | 2.06051 |
| foxconn | ARIMA(1,1,0) | AR1 0.1553 | 2.297186 | 3.18774 |
| Taiwan Cement | ARIMA(1,1,0) | AR1 0.1429 | 2.162356 | 1.06368 |
| Taiwan Semiconductor | ARIMA(0,1,1) with drift | MA1 0.1094 DRIFT 0.3590 | 1.841124 | 3.36421 |

Then, the calculation with the evaluation index is shown. Fig. 4 is using Relu activation function with 1000 epochs. After performing our proposed method, we can observe that Taiwan Cement Corp.'s forecast is better, probably because the stock price is not dramatically. We have observed that Relu Activation Function with 1000 epochs achieves the best results with MSE of 1.9%, RMSE of 1.3.

**TABLE 2.** Company and Evaluation Index.

| Company | Evaluation Index | | | | | |
|---|---|---|---|---|---|---|
| | *KD* | *MACD* | *RSI* | *OBV* | *MSE* | *RMSE* |
| Foxconn | hold | hold | hold | hold | 9.3% | 3% |
| Taiwan Cement Corp. | Buy | hold | hold | sell | 1.9% | 1.3% |
| Taiwan Semiconductor Manufacturing Company | Buy | hold | hold | hold | 81% | 9% |
| Formosa Petrochemical Corporation | buy | hold | hold | sell | 73% | 8% |
| Quanta Computer Inc. | hold | hold | hold | sell | 5.8% | 2.4% |

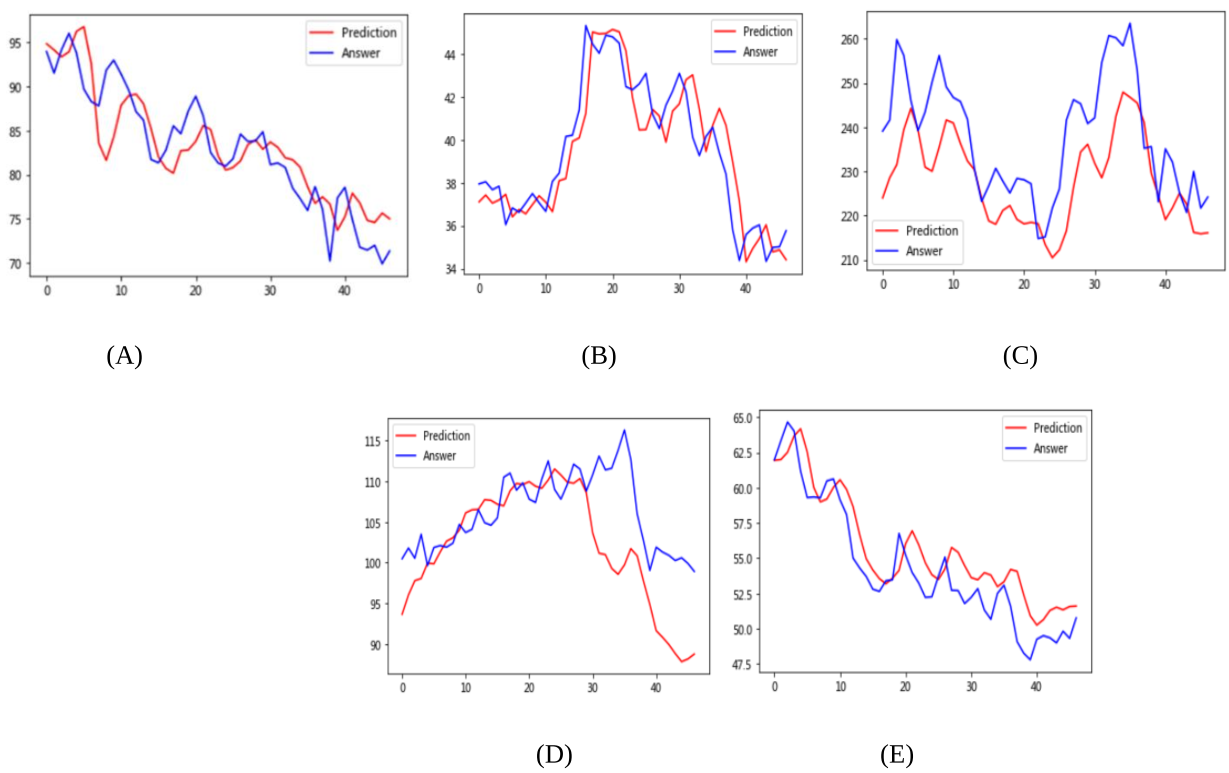


**FIGURE 4.** Prediction Foxconn (A), Taiwan Cement Corp. (B),(C) Taiwan Semiconductor, (D) Formosa Petrochemical, (E**.** Quanta Computer.

## 4. Conclusions

Stock market exchanges have become popular, encouraging researchers to find predictions using new technologies or methods. Proper predictive techniques can help researchers, investors and anyone dealing with the stock market. To help predict the stock index, a less error of the predictive model is needed which may take into account the processing of the input data. RNN cannot learn to connect information because old stored memory will be increasingly useless with time running due to overwritten or replaced new memory. Forecast using the LSTM method starts with entering inputs and outputs previously into the forget layer. Forget layer functions to produce output in the form of numbers 0 or 1 in the cell state. Number 0 means that the input will be forgotten and otherwise number 1 indicates that the input will continue. Next, at the input gate, the layer will be determined which data will be updated, and the layer will make one new candidate value. The output from the input gate layer and the layer will be combined into the cell state. Next, the old cell state will be updated with a new cell state. In contrast to RNN, LSTM does not have these deficiencies because LSTM can manage memory in each input using memory cells and gate units. Moreover, LSTM is a neural network architecture that is very powerful for processing sequential data. In this work, we use Relu activation. The Rectified Linear Unit (ReLU) will eliminate vanishing gradient by applying the element activation function as $f(x) = \max(0, x)$ and element activation will be using when it is at a threshold of 0. However, the strength using ReLU Can accelerate the stochastic gradient compared to the sigmoid function / tanh because ReLU is linear. Also. Does not use exponential operations such as sigmoid / tanh, so it can do this by creating an activation matrix when the threshold is 0.However, ReLU can be fragile during training because the large gradients flowing through ReLU cause update weights, so the neurons are inactive at the data point again. If this happens, then the gradient flowing through the unit will be zero forever from that point. That is, the ReLU units can die irreversibly during training because they can disable the manifold data.

At the same time, in this work we using long-term, short-term memory to find a new value for five days and use the predicted value as the average for the next five days, which helps investors, analysts or anyone interested in investing in the stock market. Let people understand the future of the stock market.

Moreover, use the evaluation index to consider whether to buy, sell, or hold the current state of the stock. In a nutshell, we make a comparison between traditional arima and LSTM. It was found that in the Taiwan semiconductor index following ARIMA Drift and all models had good performance as well as LSTM which was able to do learning on fluctuating time series data.

## Acknowledgments

This paper is supported by the Ministry of Science and Technology, Taiwan. The Nos are MOST-107-2221-E-324 -018 -MY2 and MOST-106-2218-E-324 -002, Taiwan. This research is also partially sponsored by Chaoyang University of technology (CYUT) and Higher Education Sprout Project, Ministry of Education, Taiwan, under the project name: “The R&D and the cultivation of talent for Health-Enhancement Products."

## References

[1] J. Zhang, Y. Zheng, and D. Qi, “Deep spatio-temporal residual networks for citywide crowd flows prediction,” in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[2] Z. Lipton, “Learning to diagnose with LSTM recurrent neural networks,” *arXiv Prepr.*, 2017.

[3] A. Grover, A. Kapoor, and E. Horvitz, “A Deep Hybrid Model for Weather Forecasting,” 2015.

[4] R. E. Caraka, S. A. Bakar, M. Tahmid, H. Yasin, and I. D. Kurniawan, “Neurocomputing Fundamental Climate Analysis,” *Telkomnika*, vol. 17, no. 4, 2019.

[5] A. Ghaderi, B. M. Sanandaji, and F. Ghaderi, “Deep Forecast: Deep Learning-based Spatio-Temporal Forecasting,” *arXiv Prepr.*, 2017.

[6] X. Qiu, Y. Ren, P. N. Suganthan, and G. A. J. Amaratunga, “Empirical Mode Decomposition based ensemble deep learning for load demand time series forecasting,” *Appl. Soft Comput. J.*, 2017.

[7] R. E. Caraka and S. A. Bakar, “Evaluation Performance of Hybrid Localized Multi Kernel SVR (LMKSVR) In Electrical Load Data Using 4 Different Optimizations,” *J. Eng. Appl. Sci.*, vol. 13, no. 17, 2018.

[8] N. Phan, Y. Wang, X. Wu, and D. Dou, “Differential Privacy Preservation for Deep Auto-Encoders: An Application of Human Behavior Prediction,” *Aaai*, 2016.

[9] N. Laptev, J. Yosinski, L. Erran, and L. Slawek, “Time-series Extreme Event Forecasting with Neural Networks at Uber,” in *International Conference on Machine Learning*, 2017.

[10] T. Toharudin, R. E. Caraka, A. Bachrudin, S. A. Bakar, and D. Ambarsari, “Copula-Based Feedforward Neural Network Genetic Algorithm Cargo Forecasting,” *Int. J. Adv. Soft Comput. Its Appl.*, vol. 11, no. 1, pp. 22–23, 2019.

[11] Y. Tian, K. Zhang, J. Li, X. Lin, and B. Yang, “LSTM-based traffic flow prediction with missing data,” *Neurocomputing*, 2018.

[12] G. S. Atsalakis and K. P. Valavanis, “Surveying stock market forecasting techniques - Part II: Soft computing methods,” *Expert Syst. Appl.*, 2009.

[13] J. L. Elman, “Finding structure in time,” *Cogn. Sci.*, 1990.

[14] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput.*, 1997.