

Comparison of Feedforward Neural Network with Different Training Algorithms for Bitcoin Price Forecasting

Eng Chuen Loh¹, Shuhaida Ismail^{1*}, Azme Khamis¹ and Aida Mustapha²

¹*Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, Pagoh Edu Hub, 84600, Muar, Johor, Malaysia.*

²*Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, 86400, Batu Pahat, Johor, Malaysia.*

Bitcoin is the most popular cryptocurrency with the highest market value. It was said to have potential in changing the way of trading in future. However, Bitcoin price prediction is a hard task and difficult for investors to make decision. This is caused by nonlinearity property of the Bitcoin price. Hence, a better forecasting method are essential to minimize the risk from inaccuracy decision. The aim of this paper is to compare two different training algorithms which are Levenberg-Marquardt (LM) backpropagation algorithm and Scaled Conjugate Gradient (SCG) backpropagation algorithm using Feedforward Neural Network (FNN) to forecast the Bitcoin price. After obtaining the forecasting result, forecast accuracy measurement will be carried out to identify the best model to forecast Bitcoin price. The result showed that the performance of Bitcoin price forecasting increased after the application of FNN – LM model. It is proven that Levenberg-Marquardt backpropagation algorithm is better compared to Scaled Conjugate Gradient backpropagation when forecasting Bitcoin price using FNN. The resulting model provides new insights into Bitcoin forecasting using FNN – LM model which directly benefits the investors and economists in lowering the risk of making wrong decision when it comes to invest in Bitcoin.

Keywords: Bitcoin Price; Artificial Neural Network; Forecasting

I. INTRODUCTION

Cryptocurrency has now been the hot topic worldwide. It is said to be the revolution of the currency transaction (Almeida *et. al.*, 2015). Bitcoin was first introduced by a pseudonym creator, Satoshi Nakamoto in the late 2008 in the internet (Nakamoto, 2008). As Bitcoin price did not rely on any government bodies or agencies, therefore, Bitcoin has the feature of decentralized system. This feature has influenced the demand of Bitcoin in the market. Moreover, this currency is built based on advance mathematics and computer engineering, therefore, it is difficult to counterfeit the currency and the user are safe from cyber-attack which might cause losses or leakage of identity (Gervais *et. al.*, 2014). Blockchain has been used as Bitcoin's master ledger. From

Blockchain record, the Bitcoin market capitalization has increase to 280 billion USD in the end of 2017, which is an increase of approximately 10000% from the early 2012 (Thies *et al.*, 2018). The increasing trend of Bitcoin price resulting in increasing the attention from investors and economists around the world.

As Bitcoin has becoming one of the major contributors to the cryptocurrency market. It plays an important role in developing a brand-new decentralized currency transactions system while influencing the stock market direction. However, Bitcoin prices are frequently fluctuating because of its virtual characteristics. Therefore, it is essential to develop a reliable model in order to enable investors and economists to forecast the Bitcoin price so that they can make correct and rational decision to invest in Bitcoin at the optimum time and

*Corresponding author's e-mail: shuhaida@uthm.edu.my

price. Bitcoin price has complex and nonlinearity characteristics (Polasik *et. al.*, 2015; Zhou *et. al.*, 2017). Thus, classical forecasting methods such as Holt-Winters is not suitable to forecast Bitcoin price data because the methods require the assumption of linearity (Mcnally, 2016). Therefore, a method which able to handle nonlinearity and complexity of the data is needed to forecast Bitcoin price.

With the development of modern technology, Artificial Intelligence (AI) is one of the main focus of development and Artificial Neural Network (ANN) is one of the AI forecasting method (Wang *et. al.*, 2009). ANN was inspired by the characteristic of human brain neuron that mimics the way human brain processes information (Beresford & Agatonovic, 2000). The system consists of nodes that interconnected with each other by different layers of different tasks (Sharma *et. al.*, 2012). ANN able to learn and improve over time by considering examples and information of the given data sets. ANN is widely used in different field including price prediction, voice recognition, pattern recognition, search engine filtering, gaming, medical diagnosis and information processing. This is because ANN is able to handle highly nonlinear and complex situation (Molaie *et. al.*, 2014). Thus, many researchers starting to apply ANN method in forecasting time series.

This study aims to forecast the future values of Bitcoin price using FNN with two different training algorithms which are Levenberg-Marquardt backpropagation algorithm and Scaled Conjugate Gradient backpropagation algorithm. Levenberg-Marquardt backpropagation algorithm was chosen as one of training algorithm in this research because of it is commonly used in training the FNN whereas Scaled Conjugate Gradient backpropagation algorithm was chosen because it able to perform significantly when dealing with large network or large dataset (Chel, *et. al.*, 2011). Besides, both of these algorithms avoid the Hessian matrix which indirectly increase the training speed (Kişi & Uncuoğlu, 2005). The performance of each algorithm will be evaluated using several statistical performances and the best model will be selected accordingly.

II. METHODOLOGY

This section discusses in detail about FNN and data pre-processing technique used for Bitcoin price forecasting.

A. Statistical Testing

In order to determine the characteristics and nature of the data, two statistical tests are carried out which are Anderson Darling (AD) Test and Augmented Dickey Fuller (ADF) Test. AD Test is used to determine the normality of the dataset (Razali & Wah, 2011). The hypotheses of this test are as follows:

H_0 : The data follows the normal distribution

H_1 : The data does not follow the normal distribution

The p -value obtained is then compared with the critical p -value of 0.05 and reject the null hypothesis when the p -value smaller than the critical p -value. On the other hand, ADF Test is used to identify the stationarity of a data set (Mushtaq, 2011). The hypotheses of this test are as follows:

H_0 : The data are not stationary

H_1 : The data are stationary

The p -value obtained is then compared with the critical p -value of 0.05 and reject the null hypothesis when the p -value smaller than the critical p -value.

B. Feedforward Neural Network

Feedforward Neural Network is one of the basic forms of Artificial Neural Network which pass the information from the input layer directly to the output layer after undergoing activation function (Rosenblatt, 1961). However, in order to handle nonlinear data, a hidden layer is needed to be insert within the input and output layer (Negnevitsky, 2003). Figure 1 illustrates the model of the FNN.

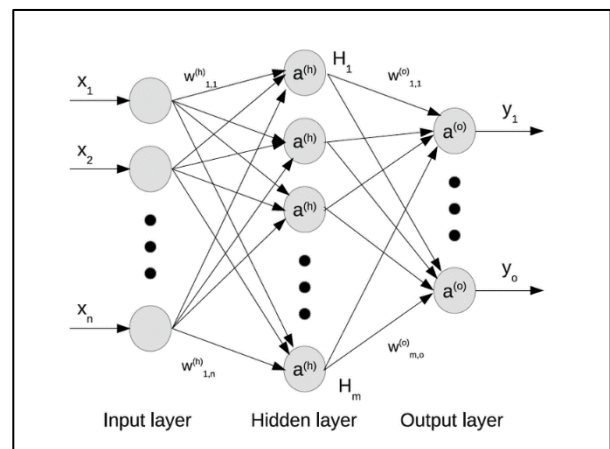


Figure 1. Architecture of FNN (Parvizi *et. al.*, 2014)

The weight for each of the interconnection constantly changes based on the predetermined training algorithms. In this research, there are two different training algorithms are used to forecast the Bitcoin price data which are Levenberg-Marquardt backpropagation algorithm and Scaled Conjugate Gradient backpropagation algorithm. Levenberg-Marquardt algorithm is modified from Gauss Newton method (Reynaldi *et. al.*, 2012). The modification is as shown in Equation (1).

$$w_{n+1} = w(n) - \left(J(n)^T J(n) + \mu I \right)^{-1} J(n)^T e(n) \quad (1)$$

where w is the weight of the connection ;If error decreased, μ equals to μ/θ and else if error increase, μ equals to $\mu \times \theta$ where matrix inverse will always produce a result; $\theta = 0.01 (e^t e)$; e = Error matrix; J = Jacobi matrix; I is the Identity matrix.

Scaled Conjugate Gradient back propagation algorithm is the combination of Levenberg-Marquardt algorithm with Conjugate Gradient approach (MØller, 1993). The algorithm is as shown in Equation (2).

$$\tilde{w}_k = \tilde{s}_k + \left(\bar{\lambda}_k - \lambda_k \right) \tilde{p}_k \quad (2)$$

with

$$\bar{\lambda}_k = 2 \left(\lambda_k - \frac{\delta_k}{\left(\tilde{p}_k \right)^2} \right) \quad (3)$$

$$\delta_k = \tilde{p}_k^T \tilde{s}_k \quad (4)$$

$$\tilde{s}_k = \frac{E'(\tilde{w}_k + \sigma_k \tilde{p}_k) - E'(\tilde{w}_k)}{\sigma_k} \quad (5)$$

where \tilde{w}_k is the weight at period k ; \tilde{s}_k is the step size at period k ; λ_k is the scalar in conjugate gradient that regulate the indefiniteness of $E''(\tilde{w}_k)$; \tilde{p}_k is the conjugate weight vector at period k ; δ_k is the changes of the weight at period k ; σ_k is the weight's standard deviation at period k .

The predicted output values are obtained from the Equation (6) after obtaining the weight for each connection between nodes (Ismail *et al.*, 2011).

$$y_i = f \left[\sum_{j=1}^{\beta} w_j g \left(\sum_{i=1}^{\alpha} w_i x_i + w_i \theta \right) + w_j \theta \right] \quad (6)$$

with

$$g(x) = \tan \operatorname{sig}(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (7)$$

$$f(x) = \operatorname{purelin}(x) = x, \quad (8)$$

Where y_i is the output value at actual time t ; x_i is the input value at actual time t ; w_i is the connection weight between input and hidden layer nodes; w_j is the connection weight between hidden and output layer nodes; θ is the bias constant; $f(x)$ is the activation function from hidden layer to output layer; $g(x)$ is the activation function from input layer to output layer; α is the number of input nodes; β is the number of hidden nodes; $\tan \operatorname{sig}(x)$ represents the tan-sigmoid activation function while $\operatorname{pureline}(x)$ represents the pure-linear activation function.

C. Training Algorithm Parameter

The parameter shown in Table 1 were used for Levenberg-Marquardt and Scaled Conjugate Gradient training algorithms respectively.

Table 1. Parameter for Each Models

Parameter	LM	SCG
Maximum epochs	10000	10000
Minimum gradient	1.00×10^{-6}	1.00×10^{-6}
Maximum fail	50	50
Learning rate, α	0.1	0.1
Performance goal	0	0
Training gain, μ	1.00×10^{-3}	-
Maximum μ	1.00×10^{10}	-

Both Levenberg-Marquardt and Scaled Conjugate Gradient training algorithms are set with the same parameter. The training stops when the algorithm satisfied several conditions, where the process reached maximum epochs of 10000 iterations, minimum gradient of 1.00×10^{-6} unit, maximum fail in validation of 50 times, learning rate of 0.1 unit and the performance goal of 0. Aside for Scaled Conjugate Gradient training algorithm, Levenberg-Marquardt training algorithm was set with μ of 1.00×10^{-3} unit and a maximum μ of 1.00×10^{10} . The numbers of iteration and maximum fail in validation is set to such amount in order to let the model to train more vividly compared to the lower amount of the standard numbers of iteration and maximum fail in validation. The other parameters are set to be default according to the Matlab's preset value.

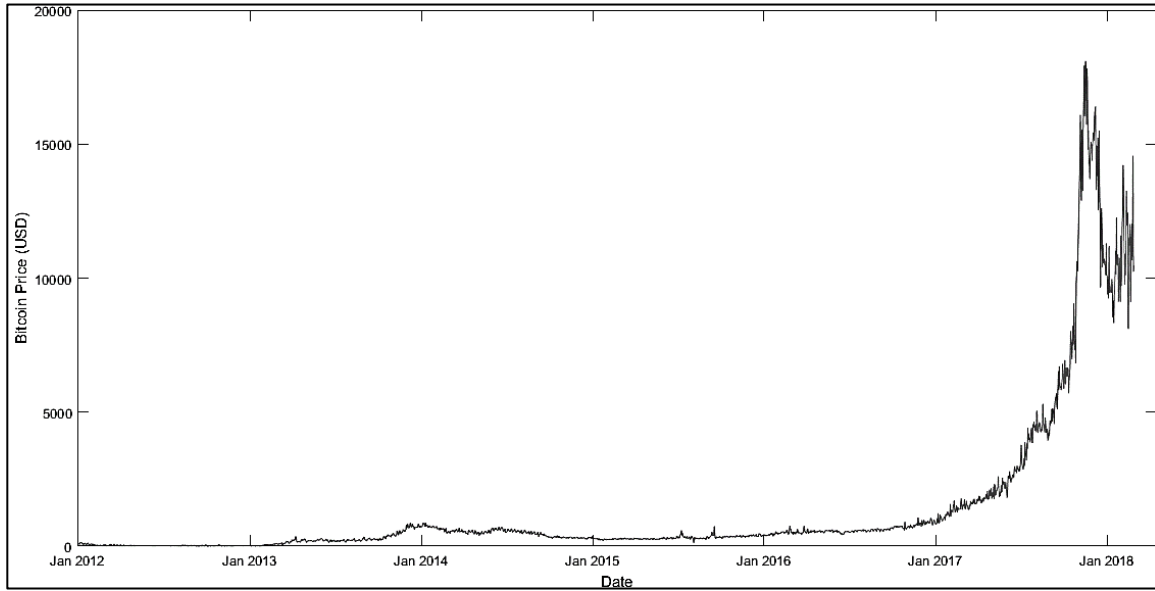


Figure 2. Times Series Plot

D. Data Pre-processing

Different value range in the variables will directly influence the tendency and accuracy for the models especially for neural network (Zhang, 2012). Therefore, normalization method is applied in the analysis. The data used in the analysis are transformed using Min-Max normalization method which transforms the data into a defined range of 0 to 1 (Chou *et. al.*, 2016). The equation of the Min-Max normalization method is shown in Equation (9).

$$x_i = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (9)$$

Where x_i is the normalized x value at time i ; x is the actual x value; x_{min} is the minimum actual value; x_{max} is the maximum actual value.

E. Forecast Accuracy

In this research, five types of forecast accuracy measurements are applied to evaluate the accuracy of the predicted output for the forecasting models. The measurements that will used in this study are Mean Absolute Error (MAE), Mean Forecast Error (MFE), and Root Mean Square Error (RMSE) (Hyndman & Koehler, 2006). Whereby best model will be selected based on the smallest values for all measurements (Lee *et. al.*, 2012). This measurement is used to determine how the errors of the model distribute and magnitude of them.

The equation for each of the forecast accuracy are shown as follows:

$$MAE = \frac{\sum_{i=1}^n |y_t - \hat{y}_t|}{n} \quad (10)$$

$$MFE = \frac{\sum_{i=1}^n (y_t - \hat{y}_t)}{n} \quad (11)$$

$$RMSE = \frac{1}{n^2} \sqrt{\sum_{i=1}^n (y_t - \hat{y}_t)^2} \quad (12)$$

where y_t is the actual values at time t ; \hat{y}_t is the predicted values at time t ; n is the number of observations.

F. Data

Different websites offer different selling price for Bitcoin. For this research, the Bitcoin price data were collected from Blockchain, which is the master ledger that records the original Bitcoin price (Quandl, 2018). There were 2251 observations of daily Bitcoin price data starting from 1st January 2012 until 28th February 2018 were used in this study as shown in Figure 2. Aside from Bitcoin price data, there were several others daily data variables collected from Blockchain which includes hash rate, average block size, transaction cost, number of transactions, miner revenue and

number of transactions per block. These are some of the influence factor mentioned by Kristoufek (2015).

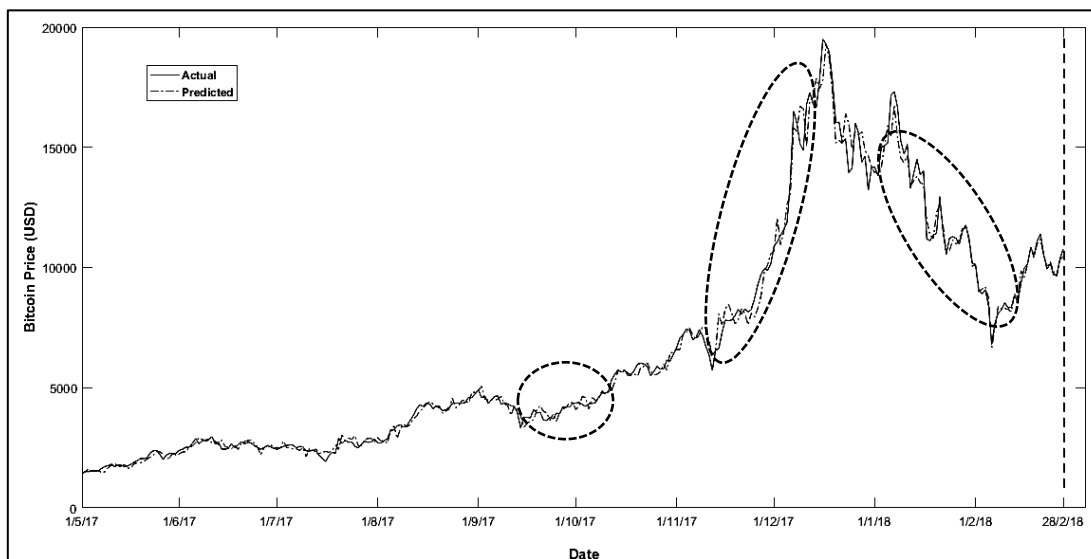


Figure 3. Actual versus Predicted Bitcoin Price for Last 10 Months (LM Algorithm)

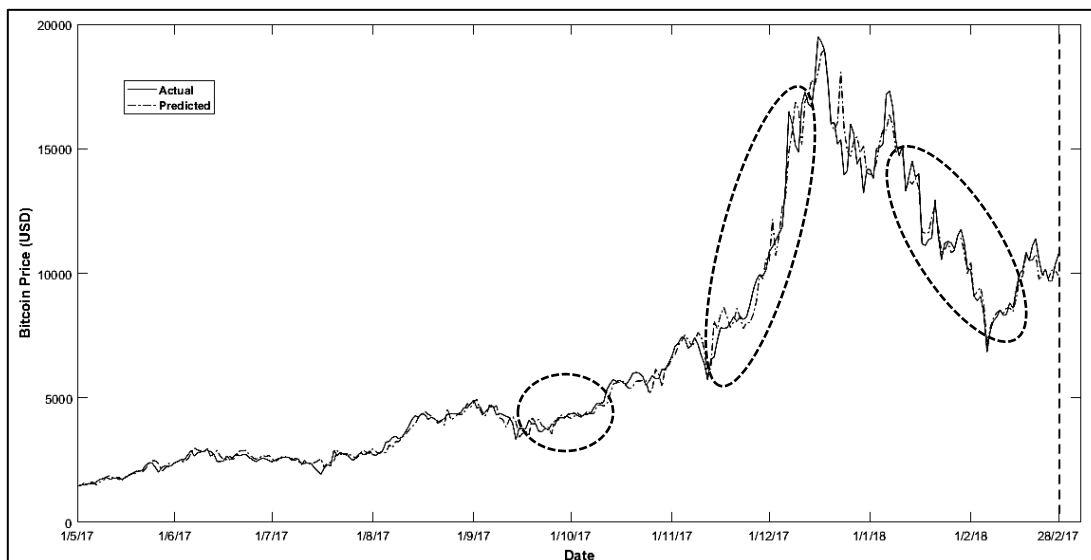


Figure 4. Actual versus Predicted Bitcoin Price for Last 10 Months (SCG Algorithm)

III. RESULTS AND DISCUSSIONS

Based on visual inspection on Figure 2, the series has showed to be nonlinear and non-stationary. However, proper statistical tests needed to carry out to prove the findings. Therefore, Augmented Dickey Fuller (ADF) and Anderson Darling (AD) statistical tests were performed to prove the stationarity and linearity properties in the data. Figure 2 shows the output for ADF test. The result showed a logical result of value 0 with the p -value of 0.8401 ($p > 0.05$). It indicates that this test fails to reject the null hypothesis of a

unit root is equal to 0, suggesting that the data is not stationary. Meanwhile, the obtained output for AD test shows a logical result of value 1 with p -value approximate to 0.0005 which indicates that null hypothesis is rejected. Thus, it can be concluded that the data does not follows the normal distribution, hence nonlinearity does exist in the data. Evidences from AD and ADF test indicate that Bitcoin time series data is proven to be non-stationary and nonlinear, hence, fulfil the assumption of Neural Network.

Once the non-stationarity and nonlinear properties proved in the dataset, the dataset is ready to be used in predication

by using NN method. In this study, the NN model will be using two different algorithms which are Levenberg-Marquardt backpropagation and Scaled Conjugate Gradient backpropagation. The predicted values were computed using forecast accuracy which are MAE, MFE and RMSE. The forecast accuracy of the models is then tabulated in Table 2.

Table 2. Forecast Accuracy of the Models

Models	Forecast Accuracy		
	MAE	MFE	RMSE
FNN – LM	63.718	-0.98	160.729
FNN – SCG	72.030	-0.69	175.439

From Table 2, it is apparent that FNN with LM algorithm was chosen as the best model. This is because FNN-LM has the smallest MAE and RMSE values compared to FNN-SG, indicating that FNN-LM produced higher accuracy in forecasted results. However, FNN-SG has over-forecast the future Bitcoin price by based on MFE results, FNN-LM and FNN-SCG showed that both algorithms has slightly over-forecast the future value of Bitcoin price with an average absolute error of 63.718 units and 72.030 units respectively.

According to Figure 3 and Figure 4, there are several points that can briefly explain how well the predicted value for each model fits to the actual value. The first point was in October 2017 whereby Figure 3 shows an under-forecast trend compared to the predicted values in Figure 3. When the Bitcoin price started to increase drastically at the end of 2017, both models able to capture the trend of steep increment. However, Bitcoin price begins dropping early 2018 due to falling domino effect. In that event, FNN-SCG model has over-forecasted the Bitcoin price. Interestingly, FNN-LM was proven able to produce predicted values that very close to the actual observations. Although Bitcoin price started to drop

deep down in February 2018, FNN-LM model still able to predict accurately compared to FNN-SCG that shown in Figure 3.

IV. CONCLUSION

The ADF and AD tests indicate Bitcoin price has the characteristics of nonlinear and non-stationary. Therefore, classical forecasting methods are not suitable to forecast Bitcoin price as the classical forecasting methods require to fulfil the linearity and stationary assumption. Feedforward Neural Network with Levenberg-Marquardt backpropagation algorithm has the lowest error compared to Neural Network with Scaled Conjugate Gradient backpropagation algorithms in MAE, MFE and RMSE, with the values of 63.718, 32.764%, and 160.729 respectively. Thus, Levenberg-Marquardt is the best backpropagation methods to be applied in Feedforward Neural Network when forecasting Bitcoin Price.

V. ACKNOWLEDGEMENTS

This research was supported by the Universiti Tun Hussein Onn Malaysia under the Tier 1 Grant Scheme Vot H073 and Ministry of Education Malaysia under the Fundamental Research Grant Scheme (FRGS) Vot K082.

VI. REFERENCES

Almeida, J, Tata, S, Moser, A & Smit, V 2015, 'Bitcoin prediction using ANN', *Neural Network*

Beresford, R & Agatonovic, KS 2000, 'Basic concept of Artificial Neural Network (ANN) modelling and its application in pharmaceutical research', *Journal of Pharmaceutical and Biomedical Analysis*, vol. 22, no. 5, pp. 717 – 727.

Chel, H, Majumder, A, & Nandi, D 2011, 'Scaled Conjugate Gradient algorithm in Neural Network based approach for handwritten text recognition'. *Trends in Computer Science, Engineering and Information Technology*, Berlin, Heidelberg: Springer, pp. 196 – 210.

Chou, JS & Thedja, JPP 2016, 'Metaheuristic optimization within machine learning- based classification system for

- early warning related to geotechnical problems', *Automation in Construction*, vol. 68, pp. 65 – 80.
- Gervais, A, Karame, GO, Capkun, V & Capkun, S 2014, 'Is Bitcoin a decentralized currency?', *IEEE Security and Privacy*, vol. 12, no. 3, pp. 54 – 60.
- Hyndman, RJ & Koehler, AB 2006, 'Another look at measures of forecast accuracy', *International Journal of Forecasting*, vol. 22, no. 4, pp. 679 – 688.
- Ismail, Z & Khamis, A 2011, 'Neural network in modeling Malaysian oil palm yield', *American Journal of Applied Science*, vol. 8, no. 8, pp. 796 – 803.
- Kişi, Ö & Uncuoğlu, E 2005, 'Comparison of three back-propagation training algorithms for two case studies'. *Indian Journal of Engineering and Materials Sciences*, vol. 12, no. 5, pp. 434 – 442.
- Kristoufek, L 2015, 'What are the main drivers of Bitcoin price? Evidence from Wavelet Coherence Analysis. *PloS ONE*, vol. 10, no. 4, pp. 1 – 15.
- Lee, MH, Nor, ME, Sadaei, HJ, Rahman, NHA & Kamisan, NAB 2012, 'Fuzzy time series: An application to tourism demand forecasting', *American Journal of Applied Sciences*, vol. 9, no. 1, pp. 132 – 140.
- McNally, S 2016, 'Predicting the price of Bitcoin using machine learning', National College of Ireland: MSc Research Project.
- Molaie, M, Falahian, R, Gharibzadeh, S, Jafari, S & Sprott, JC 2014, 'Artificial Neural Networks: Powerful tools for modeling chaotic behavior in the nervous system', *Frontiers in Computational Neuroscience*, vol. 8, pp. 2013 – 2015.
- MØller, MF 1993, 'A scaled conjugate gradient algorithm for fast supervised learning', *Neural Networks*, vol. 6, pp. 525 – 53.
- Mushtaq, R 2011, 'Testing time series data for stationarity', retrieved on February 15, 2018, from <http://ssrn.com/abstract=19110681>.
- Nakamoto, S 2008, 'Bitcoin: A peer-to-peer electronic cash system', available at: <https://bitcoin.org> [Assessed 6 May 2018].
- Negnevitsky, M 2003, *Artificial Intelligence*, 2nd ed., England: Addison Wesley.
- Parvizi, M, Paegert, M & Stassun, KG 2014, 'The EB factory project. II. Validation with the Kepler field in preparation for K2 and TESS', *Astronomical Journal*, vol. 148, no. 6, pp. 125.
- Polasik, M, Piotrowska, A & Wisniewski, TP 2015, 'Price fluctuations and the use of Bitcoin: An empirical inquiry', *International Journal of Electronic Commerce*, vol. 20, no. 1, pp. 9 – 49.
- Quandl 2018, 'Bitcoin market price USD', retrieved on March 15, 2018 from <https://www.quandl.com/data/BCHAIN/MKPRU-Bitcoin-Market-Price-USD>
- Razali, NM & Wah, YB 2011, 'Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling test', *Journal of Statistical Modeling and Analytics*, vol. 2, no. 1, pp. 21 – 33.
- Reynaldi, A, Lukas, S & Margaretha, H 2012, 'Backpropagation and Levenberg-Marquardt algorithm for training finite element neural network', in AMSS 6th European Modelling Symposium (EMS 2012), Malta: UK Sim, pp. 89 – 94.
- Rosenblatt, F 1961, *Principle of Neurodynamics. Perceptrons and the Theory of Brain Mechanism*, Buffalo, New York: Cornell Aeronautical Lab Inc.
- Sharma, V, Dev, A & Rai, S 2012, 'A comprehensive study of artificial neural network', *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 2, no. 10, pp. 278 – 284.
- Thies, S & Molnár, P 2018, 'Bayesian change point analysis of Bitcoin returns', *Finance Research Letter*, available at: <https://www.sciencedirect.com/science/article/abs/pii/S1544612318300710>.
- Wang, WC, Chau, KW, Cheng, CT & Qiu, L 2009, 'A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series', *Journal of Hydrology*, vol. 374, no. 3, pp. 294 – 306.
- Zhang, GP, 2012, 'Neural network for time-series forecasting', eds G Rozenberg, T Bäck & JN Kok, in *Handbook of Natural Computing*, Berlin, Heidelberg: Springer, pp. 461 – 477.
- Zhou, Q, Zhang, Q & Zhang, QI 2017, 'Agent-based simulation research on Bitcoin price fluctuation', in 2nd International Conference on Artificial Intelligence and Engineering Application (AIEA 2017), Guilin, China: DEStech Publications Inc., pp. 21 – 33.