Cryptocurrency Price Prediction with Neural Networks of LSTM and Bayesian Optimization

Ehsan Sadeghi Pour, Hossein Jafari, Ali Lashgari, Elahe Rabiee, and Amin Ahmadisharaf

ABSTRACT

In this paper we present a price prediction for Bitcoin prices. The methodology used is a hybrid artificial neural network model of Long Short-Term Memory and Bayesian Optimization. This is a complex model with a high prediction power, which to our knowledge has not been applied to prediction of cryptocurrency prices to date. Following Charandabi and Kamyar (2021A), we elaborate on previous methods used for prediction of cryptocurrency prices and build on their methodology. We conclude with detailed graphs and tables of optimization results.

Keywords: Bayesian optimization, artificial neural networks, cryptocurrency price prediction, long short-term memory.

I. INTRODUCTION

The world has gone through many changes since 2009. Just over the last 13 years Instagram was founded, Steve Jobs passed away, a whole pandemic started, and so on. But an extremely important event that happened in 2009 has been changing the world since is the launch of Bitcoin. While the cryptocurrency was conceptualized back in 1990’s, it was not relevant to the public until years after Bitcoin was first implemented. The rate of increase in popularity is so high that Bitcoins which barely sufficed to by a pizza in 2010, costed as high as $68000 in 2021.

Financial analysts attribute the popularity of Bitcoin to its decentralization, secure blockchain system, and control over value. Charandabi and Kamyar (2021A) presents a thorough survey of the history, foundation, and relevance of Bitcoin. In the same paper, they argue on the importance and relevance of artificial neural networks to predict prices of cryptocurrency. Artificial neural networks may have a high predictive power and can be pertained to cryptocurrency price data in any time horizon.

Artificial neural network models can be used in combination with one another to improve accuracy and predictive power. Charandabi and Kamyar (2021B) presents a survey on the literature on applications of artificial neural networks to cryptocurrency price prediction. Based on their reports, there have been a number of hybrid and singular artificial neural network models implemented so far; however, there are still numerous gaps to be filled in the literature.

This paper serves to fill the literature by providing a model of Neural Networks of LSTM and Bayesian Optimization. In Section II, we review related literature; in Section III, we discuss the data process, in Section IV we go through optimization details, and in section V we conclude.

II. LITERATURE REVIEW

“Artificial neural networks (ANNs) are biologically inspired computational networks. Multilayer perceptrons (MLPs), the ANNs most commonly used for a wide variety of problems, are based on a supervised procedure and comprise three layers: input, hidden, and output.” (Park & Lek, 2016) In this context, artificial neural networks may be...
view the data aspect of this research paper as a novel work in the literature. All exchange rates are with respect to United States Dollars.

IV. OPTIMIZATION

The baseline treatment is prediction of Bitcoin prices with Neural Networks of LSTM and Bayesian Optimization. The value of optimal delays is set to 1.30, and the value of optimal training percentage (division of data to training and testing) is set to 0.8 (Nejatian, 2022). Furthermore, to optimize speed, the optimal code execution environment is the CPU. The optimal drop out value is 0.5. These values are exogenous to the model and remain constant across the treatments.

Some general deep learning parameters are exogenously varied across treatments. In the baseline, the maximum number of training Epochs in deep learning algorithms (maxEpoch) is set equal to 400. Following standard cryptocurrency optimization literature (Charandabi & Kamyar, 2021C) we exogenously set it to 200 in the treatment run. Additionally, in the baseline, the minimum batch size of training Epochs in deep learning algorithms is set equal to 32. Following standard cryptocurrency optimization literature (Charandabi & Kamyar, 2021C) we exogenously set it to 16 in the treatment run.

We start reporting results by the baseline of Bitcoin prices (weekly from January 2020 through January 2022), maximum training Epoch number of 400, and batch size of training Epochs in deep learning algorithms of 32. Fig. 1 depicts the input data, reporting the values for mean and standard deviation as well. The algorithm normalizes input data prior to running, in order for the base numbers to be feasible for computational concerns. Fig. 2 depicts the normalized input data, reporting the values for mean and standard deviation as well. In the green and blue lines, Fig. 3 depicts the minimum observed objective and estimated minimum objective, respectively.

III. DATA

We extracted data from Yahoo! Finance, a major resource for daily data on financial indices and cryptocurrency prices. Ghashami et al. (2021) elaborates on data gathering for financial indicators for the purposes of time-series prediction through artificial neural network methodology. We follow the same system for optimization of data points among the daily values for Open, High, Low, Close, and Adj. Close.

Cryptocurrency data is extremely volatile by nature. There’s a body of literature dedicated to prediction of volatility of cryptocurrency price data, as explained thoroughly in the survey paper Charandabi & Kamyar (2021C). In order to avoid running into problems and retrieving weak predictions, we use weekly data, and expand the time horizon instead. The employed data runs from January 10, 2020, to January 10, 2022, on a weekly basis. Most of the literature on prediction of cryptocurrency prices employ data from a short time horizon, as explained in the survey paper Charandabi & Kamyar (2021B). Therefore, we have selected the data for a period of two years to ensure that the model is capable of making predictions for a long time horizon.

Fig. 1. Input data of first treatment.
Fig. 2. Normalized input data of first treatment.

Fig. 3. Minimum observed objective and estimated minimum objective values.

Fig. 4-7 depict technical aspects of training data. Fig. 4 shows the output evaluation and reports the rank correlation number, which is significantly high. Fig. 5 shows the error evaluation and reports the MSE (Mean Squared Error), RMSE (Rooted Mean Square Error), and NRMSE (Normalized Rooted Mean Square Error). Figure 6 shows the error histogram, reporting Error Mean and Error Standard Deviation numbers. Fig. 7 shows the Regression Graph Evaluation of train data, reporting the R-squared value, that is highly significant.

Fig. 4. Output evaluation.

The output results had Maximum Objective Evaluations of 60 reached. Total time elapsed was 849.8 seconds, with a total function evaluation count of 60 and a total objective function evaluation time of 766.6 seconds. Observed objective function value was 0.07662, with an estimated objective function value of 0.15971 and a function evaluation time of 7.8227. Also, estimated objective function value was 0.14654 and estimated function evaluation time was 6.7355. Fig. 8-10 represent technical aspects of the test data. Fig. 8 depicts output evaluation, and reports rank correlation value. Fig. 9 shows the error evaluation and reports the MSE (Mean Squared Error).
Squared Error), RMSE (Rooted Mean Square Error), and NRMSE (Normalized Rooted Mean Square Error). Fig. 10 shows the error histogram, reporting Error Mean and Error Standard Deviation numbers.

![Error Histogram](image)

Fig. 10. Error histogram of test data.

MSE (Mean Squared Error), RMSE (Rooted Mean Square Error), and NRMSE (Normalized Rooted Mean Square Error) of all data. Fig. 12 depicts regression evaluation of all data, reporting the R-squared value.

![Regression Graph](image)

Fig. 12. Regression graph evaluation.

Table 1 shows technical details of the best observed feasible point (above) and the best estimated feasible point (below), according to the models. All points are on the first layer.

| Number of Units | LSTM Layer | Initial Learn Rate | Layer 2 Reg. |
|-----------------|------------|--------------------|--------------|
| 68              | 2          | 0.016              | 0.00475      |
| 71              | 2          | 0.021              | 0.00798      |

The treatment run employed the same methodology and data, with the exception of the maximum number of training Epochs in deep learning algorithms (maxEpoch) set equal to 200, and the minimum batch size of training Epochs in deep learning algorithms set equal to 16. In the green and blue lines, Fig. 13 depicts the minimum observed objective and estimated minimum objective, respectively.
Fig. 14-17 depict technical aspects of training data. Fig. 14 shows the output evaluation and reports the rank correlation number, which is significantly high. Fig. 15 shows the error evaluation and reports the MSE (Mean Squared Error), RMSE (Rooted Mean Square Error), and NRMSE (Normalized Rooted Mean Square Error). Fig. 16 shows the error histogram, reporting Error Mean and Error Standard Deviation numbers. Fig. 17 shows the Regression Graph Evaluation of train data, reporting the R-squared value, that is highly significant.

The output results had Maximum Objective Evaluations of 60 reached. Total time elapsed was 849.8 seconds, with a total function evaluation count of 60 and a total objective function evaluation time of 1271.6 seconds. Observed objective function value was 0.08878, with an estimated objective function value of 0.13354 and a function evaluation time of 18.863. Also, estimated objective function value was 0.071925 and estimated function evaluation time was 9.4717. Fig. 18-20 represent technical aspects of the test data. Fig. 18 depicts output evaluation, and reports rank correlation value. Fig. 19 shows the error evaluation and reports the MSE (Mean Squared Error), RMSE (Rooted Mean Square Error), and NRMSE (Normalized Rooted Mean Square Error). Fig. 20 shows the error histogram, reporting Error Mean and Error Standard Deviation numbers. Fig. 21 and 22 show technical details of all (i.e., train and test) data. Fig. 21 depicts the error evaluation and reports the MSE (Mean Squared Error), RMSE (Rooted Mean Square Error), and NRMSE (Normalized Rooted Mean Square Error) of all data. Fig. 22 depicts regression evaluation of all data, reporting the R-squared value.

Table II shows technical details of the best observed feasible point (above) and the best estimated feasible point (below), according to the models.
TABLE II: BEST FEASIBLE POINTS

| Number of Layer | Number of Units | LSTM Layer | Initial Learn Rate | Layer 2 Reg. |
|-----------------|-----------------|------------|--------------------|--------------|
| 3               | 61              | 1          | 0.021              | 0.00881      |
| 1               | 54              | 2          | 0.012              | 0.00889      |

While the 200-16 model provides less noise in the resulting data and filters volatilities, the 400-32 model yields a shorter elapsed time and higher accuracy factors.

Fig. 18. Output evaluation of test data.

Fig. 19. Error mean evaluation of test data.

Fig. 20. Error histogram of test data.

V. CONCLUSION

In this paper, we applied a novel algorithm for prediction of time-series data (through hybrid artificial neural networks of Long Short-Term Memory and Bayesian Optimization) to prediction of Bitcoin data. We ran two treatments of data and training variables using Bitcoin prices (weekly from January 2020 through January 2022): maximum training Epoch number of 400 and 200, and batch size of training Epochs in deep learning algorithms of 32 and 16. Results were reported graphically and in tables, and optimal solutions were comparatively shown.

There exist many possible extensions to this paper. One may run the same algorithm with different data to compare consistency of external validity to other contexts. Other bitcoin prices (e.g., Ethereum) may be good candidates. Alternatively, within the financial data realm, stock market indices could be tested with the same algorithm. Furthermore, other hybrid models such as a hybrid of the current model with a GA-ANFIS artificial neural network algorithm a la Ghashami & Kamyar (2021) could be implemented to increase the accuracy. In light of the current impact and relevance of cryptocurrency prices, it’s essential that further models be tested.
Assessment of Iran’s Wholesale Electricity Market. *Asian J. Res. Mark.*, 3, 172–177.

Nazemi, A.; Mashayekhi, M. Modelling Welfare Loss in Iranian Electricity Market. (2014) In *Proceedings of the 37th IAEE International Conference*, Energy and the Economy, New York, USA, June 2014

Nejatian, A. (2022). Time Series Prediction with Bayesian optimization, MATLAB Central File Exchange.

Nisar, M. U., Voghoei, S., & Ramaswamy, L. (2017, June). Caching for pattern matching queries in time evolving graphs: challenges and approaches. In 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS) (pp. 2352-2357). IEEE.

Park, Y.S., & Lek, S. (2016). Artificial Neural Networks. Developments in Environmental Modelling. 123–140. Springer. https://doi.org/10.1016/b978-0-444-63623-2.00007-4.

Pourbemany, J., Zhu, Y., & Bettati, R. (2021). A Survey of Wearable Devices Pairing Based on Biometric Signals. arXiv preprint arXiv:2107.11685.

Pourbemany, J., Mirjaliy, G., Abouei, J., & Raouf, A. H. F. (2018, May). Load Balanced Ad-Hoc On-Demand Routing Based on Weighted Mean Queue Length Metric. In *Electrical Engineering (ICEE)*, Iranian Conference on (pp. 470-475). IEEE.

Sak, H., Senior, A. W., and Beaufays, F. (2014). Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In *Interspeech*, Google Research.

Ulstad, A., Mashayekhi, M. and Mechstroth, H. Work in Progress: Do Engineering Students Gain Financial Literacy Skills by Taking an Engineering Economy Course? (2018) *American Society of Engineering Education (ASEE)*, Paper ID# 21471

Yahoo! (2022). Bitcoin USD (BTC-USD) price history & historical data. Yahoo! Finance. Retrieved January 15, 2022, from https://finance.yahoo.com/quote/BTC-USD/history/.

Yousefi, A., Amidi, Y., Nazari, B., & Eden, U. T. (2020). Assessing Goodness-of-Fit in Marked Point Process Models of Neural Population Coding via Time and Rate Rescaling. *Neural Computation*, 32(11), 2145-2186.