Ether Price Prediction Using Advanced Deep Learning Models

Agis Politis, Katerina Doka and Nectarios Koziris
Computing Systems Laboratory, National Technical University of Athens
{agispolitis, katerina, nkoziris}@cslab.ece.ntua.gr

***Abstract*—Over the last years, cryptocurrencies have gained popularity as a means of exchange, but mostly as an investment asset that can yield important earnings. Accurate cryptocurrency price prediction is the holy grail of investors, yet the task is extremely complex and tedious since cryptocurrencies exhibit high volatility and steep fluctuations compared to fiat money, while they depend on a plethora of factors related to the blockchain network, market trends, social popularity and the prices of other (crypto)currencies. Thus, simple statistical methods are not able to capture the complexity of cryptocurrency exchange rate, forcing researchers to turn to advanced machine learning techniques. In this work, we present a methodology for building deep learning models to forecast the price of cryptocurrencies and apply it to the prediction of Ether price, resulting in short- and long-term forecasts that achieve an accuracy of up to 84.2%.**

I. Introduction

Cryptocurrencies have come a long way since their initial appearance in 2008 [1], gaining ever-increasing popularity and social acceptance. However, in recent years, in addition to their use as a medium of exchange, cryptocurrencies have emerged as an alternative investment asset, reaching a total market capitalization of 528 billion USD in 2020 [2]. Researchers from various scientific disciplines have extensively studied the factors that influence cryptocurrency price and have leveraged techniques from the fields of statistics, machine and deep learning in order to predict it with high accuracy [3] [4] [5] [6] [7]. However, most of these models operate as a black box without considering the contribution of each feature to the prediction accuracy.

The goal of this work is to overcome the limitations of a black-box approach and present an efficient method for selecting the features that influence cryptocurrency prices the most. We propose a framework of feature selection techniques to identify the best features for predicting the future price of Ether and then develop a set of deep learning models consisting of LSTM, GRU and TCN layers that predict both the exact price of ether as well as its direction in the short and long term (one and seven days respectively). We focused on Ether because it supports much more functionality and has a more complex nature than Bitcoin due to the way it is transferred. More specifically, Bitcoin transfers can only be performed directly between user accounts. Contrarily, Ethereum, by supporting smart contracts, allows the existence of contract accounts that can cause Ether to be transferred between them unpredictably [8].

978-0-7381-1420-0/21/$31.00 ©2021 IEEE

II. Data and Feature Engineering

We used the historical daily data of Ether price from September 16, 2018 until April 16, 2020, from cryptodata-download.com, along with a set of 13 additional features that we considered important based on our domain knowledge and their significance in previous research. The dataset consisted of features related to the market trends, the Ethereum blockchain, the social popularity of cryptocurrencies as well as technical indicators. *Volume ETH*, *Volume USD* and *Bitoin's daily price* were used because they reflect market dynamics. The *number of daily transactions* on the Ethereum network as well as *the amount* corresponding to these transactions highlight the degree of trust of users in the network, while *Daily block size* and *mining difficulty* were used as two of the most important characteristics of the Ethereum blockchain. We also constructed three technical indicators: an *SMA of 14 days*, an *EMA of 14 days*, and a *MACD index*. Finally, we added the normalized *popularity index* from Google Trends for the terms *Ethereum*, *Coinbase* and *Exodus*. To the best of our knowledge, this is the first time the effect of crypto wallet's popularity in cryptocurrency price prediction is examined.

We then performed a set of feature selection methods to reduce the dataset complexity and make our models more robust and interpretable [9]. Unlike related work, we processed the daily and weekly forecasts separately as the influence of each feature might be different in these two time frames. First, we addressed the multicollinearity problem, which lies in the existence of linearly dependent features [10], by removing the features with a Pearson correlation coefficient greater than 0.8. The relative importance of each feature based on the Mean Decrease Impurity (MDI) [11] was then calculated using both the XGBoost and Random Forest algorithms, and only features with a coefficient greater than 5% were kept. Finally, the recursive feature elimination method was applied with the Random Forest and Decision Tree algorithms [12] as estimators. We maintained all features that proved to be significant even in one of the previous feature selection stages. However, the above methods provide an estimation of feature importance [11]. The features that were found to be important were tested in real forecasting conditions by using the validation set. We concluded that both the daily and weekly forecasts are influenced by the same factors that are the technical indicators EMA and MACD, the price of Bitcoin, and the search volume index of Ethereum in Google.

TABLE I: Results of daily and weekly forecasts

Model	Regression				Classification					
	RMSE		MAPE (%)		Accuracy		Precision		Recall	
	1 day	1 week	1 day	1 week	1 day	1 week	1 day	1 week	1 day	1 week
LSTM	10.6	51.2	4.4	24.7	78.9	71.9	71.1	61.3	96.4	82.6
GRU	9.6	44.2	4.2	19.6	77.2	68.4	85.7	60.0	64.3	65.2
TCN	10.5	**36.9**	4.3	**17.3**	77.1	70.2	71.4	66.7	89.3	52.2
Hybrid LSTM-GRU	**8.6**	39.5	**3.6**	20.1	80.7	70.1	77.4	63.6	85.7	60.1
Hybrid LSTM-TCN	11.2	45.7	4.8	20.1	80.6	70.2	75.8	60.0	89.3	78.3
Hybrid GRU-TCN	10.1	40.6	4.6	18.6	78.9	73.7	90.0	63.3	64.3	82.6
Ensemble	9.1	39.6	3.7	17.4	**84.2**	**78.9**	85.2	70.4	82.1	82.6

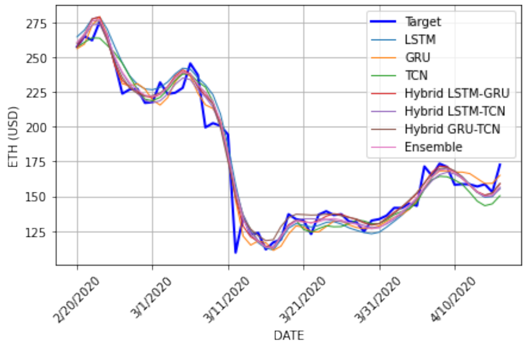

Fig. 1: Total daily predictions in the test set

Afterwards, we preprocessed the data to allow for easier pattern recognition [13], [14]. Noise reduction in a time series preserves all the important information while at the same time simplifies its form [15], so we first denoised the data using the Savitzky-Golay filter [16]. We then set the timesteps parameter which determines the amount of past information that our models rely upon to make the predictions by calculating the autocorrelation of the Ether time series [17]. The optimal number of timesteps was proved to be 30 days. The dataset was then split into train, validation and test set in proportion 80%, 10%, 10% respectively, preserving the chronological order of the data. Finally, the data were normalized between 0 and 1 by using the Min-Max scaler [18].

III. Modeling Methodology and Results

LSTM and GRU architectures are considered to be state-of-the-art in problems related to sequential data because of their ability to effectively manage past information [19]. Recently, research has shown that variations of Convolutional neural networks, like Temporal Convolutional Networks (TCN), are also very effective in sequence modeling tasks [20], achieving even better results than the LSTM and GRU architectures. To our knowledge, this is the first time that TCN models are used in time-series prediction.

We developed LSTM, GRU, and TCN models as well as hybrid models made of the above layers to make the most of the benefits that the above architectures offer. The output of each model is passed through a dense layer with one unit to produce the final output. Hybrid models consist of 2 separate models (LSTM, GRU or TCN) that have a common input and their output is concatenated before passed through the dense layer. For each model, we optimized a set of important parameters including the batch size, the learning rate, the number of layers, the size of each layer and the dropout rate at each layer using the grid search method. We also searched for the optimal optimizer considering Rmsprop and Adam and the number of epochs that we should train our models. In the case of models with TCN layers we also considered the number of filters in the convolutional layers as well as the size of the filters and the dilation values.

We evaluated the performance of the models by using the Root Mean Square Error (RMSE) and Mean Absolute Performance Error (MAPE) for the regression problem and mostly the Accuracy for the classification one. Further to the ones described above, we combined the predictions of our models creating Ensemble models. A synopsis of the results is shown in Table I and Figure 1. All models performed very well in both the regression and the classification problems. As for the daily forecasts, Hybrid LSTM-GRU model exhibited the best performance in regression making predictions with an RMSE of 8.6 and a MAPE of 3.6%. The best performance in the classification task was achieved by the Ensemble model which had an Accuracy of 84.2%. The best Ensemble model in this case was the one that combined the predictions of the LSTM, Hybrid LSTM-GRU and Hybrid LSTM-TCN models. Generally, hybrid models outperformed the individual ones and the ensemble technique led to improved results. As for the weekly forecasts, results in the regression problem were good but significantly worse than the daily ones highlighting the fact that predicting the exact price of Ether in a longer term is a very difficult problem. TCN was the best model with an RMSE of 36.9 and a MAPE of 17.3%. However, the classification results were very good, nearly as good as before. The Ensemble model had again the best performance with an Accuracy of 78.9%. This time the best Ensemble model was a combination of LSTM, GRU and Hybrid GRU-TCN.

IV. Conclusions

In this work, we describe a methodology for building accurate models to predict the exchange rate of cryptocurrencies. To that end, we propose a systematic way to identify the most appropriate data features for a specific cryptocurrency and then develop a set of state-of-the-art deep learning models for sequence prediction, including LSTM, GRU, TCN as well as model ensembles. We apply the proposed methodology to the use case of Ether: Short- and long-term forecast models predict both the exact price and the direction of Ether price, achieving an accuracy of up to 84.2%.

REFERENCES

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," Manubot, Tech. Rep., 2019.

[2] "Tradingview," https://www.tradingview.com/markets/cryptocurrencies/global-charts/, accessed: 2020-11-16.

[3] A. Greaves and B. Au, "Using the bitcoin transaction graph to predict the price of bitcoin," *No Data*, 2015.

[4] Z. Chen, C. Li, and W. Sun, "Bitcoin price prediction using machine learning: An approach to sample dimension engineering," *Journal of Computational and Applied Mathematics*, vol. 365, p. 112395, 2020.

[5] I. E. Livieris, E. Pintelas, S. Stavroyiannis, and P. Pintelas, "Ensemble deep learning models for forecasting cryptocurrency time-series," *Algorithms*, vol. 13, no. 5, p. 121, 2020.

[6] N. Smuts, "What drives cryptocurrency prices? an investigation of google trends and telegram sentiment," *ACM SIGMETRICS Performance Evaluation Review*, vol. 46, no. 3, pp. 131–134, 2019.

[7] M. Glenski, T. Weninger, and S. Volkova, "Improved forecasting of cryptocurrency price using social signals," *arXiv preprint arXiv:1907.00558*, 2019.

[8] M. Chen, N. Narwal, and M. Schultz, "Predicting price changes in ethereum," *International Journal on Computer Science and Engineering (IJCSE) ISSN*, pp. 0975–3397, 2019.

[9] "Analytics vidhya," https://www.analyticsvidhya.com/blog/2016/12/introduction-to-feature-selection-methods-with-an-example-or-how-to-select-the-right-variables/, accessed: 2020-11-16.

[10] A. Katrutsa and V. Strijov, "Comprehensive study of feature selection methods to solve multicollinearity problem according to evaluation criteria," *Expert Systems with Applications*, vol. 76, pp. 1–11, 2017.

[11] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[12] P. M. Granitto, C. Furlanello, F. Biasioli, and F. Gasperi, "Recursive feature elimination with random forest for ptr-ms analysis of agroindustrial products," *Chemometrics and Intelligent Laboratory Systems*, vol. 83, no. 2, pp. 83–90, 2006.

[13] S. Kotsiantis, D. Kanellopoulos, and P. Pintelas, "Data preprocessing for supervised leaning," *International Journal of Computer Science*, vol. 1, no. 2, pp. 111–117, 2006.

[14] S. García, J. Luengo, and F. Herrera, *Data preprocessing in data mining*. Springer, 2015.

[15] E. J. Kostelich and T. Schreiber, "Noise reduction in chaotic time-series data: A survey of common methods," *Physical Review E*, vol. 48, no. 3, p. 1752, 1993.

[16] W. H. Press and S. A. Teukolsky, "Savitzky-golay smoothing filters," *Computers in Physics*, vol. 4, no. 6, pp. 669–672, 1990.

[17] "Investopedia," https://www.investopedia.com/terms/a/autocorrelation.asp, accessed: 2020-11-16.

[18] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[19] Z. Han, J. Zhao, H. Leung, K. F. Ma, and W. Wang, "A review of deep learning models for time series prediction," *IEEE Sensors Journal*, 2019.

[20] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.