Using Machine Learning Algorithms to Predict Cryptocurrency Price

Introduction

Cryptocurrencies such as Bitcoin, despite their tremendous popularity, can still be characterized as a relatively new technology. Due to the highly volatile market prices of cryptocurrencies, combined with attractive investment opportunities, some price prediction models have been suggested (Madan et al. 1). For example, Bayesian regression analysis have been applied towards determining future price of Bitcoin and allowed to develop some profitable models. In line with such identified lack of substantial evidence, the current research aims to critically explore benefits and challenges associated with application of two machine learning algorithms – random forest and gradient boost – in relation to crypto pricing prediction, fitting, data volatility, etc. The goal of the report is to explore strengths and weaknesses of the chosen two methods in relation to cryptocurrency trading.

Critical analysis

Cryptocurrencies that have emerged in the past few years, such as Bitcoin, Etherium, Ripple, can be characterized as digital payment systems based on the development of block chain technology. For example, Bitcoin – in existence since 2009 - was based on a paper by an anonymous group calling themselves “Satoshi Nakamoto” can be considered as the first cryptocurrency. Ethereum was conceived in late in 2013 by Russian/Canadian programmer, Vitalik Buterin, and by the next year, working with Dr. Gavin Wood co-founded Ethereum.

Such currencies are decentralized, which means that they are based on peer to peer transactions and this allows the reduction of bureaucratic control over exchanges (Solanas et al. 33). Liquidity of the emerging networks are based on cryptography. Bitcoin as a thriving cryptocurrency has emerged in less than a decade, and is currently attracting a lot of attention from individual users, businessmen and investors. However, in order to ensure that such ecosystems thrive, it is important that there exist tools and mechanisms tailored towards cryptocurrencies that investors can easily use to gain benefit (Maidan et al. 1).

In relation to cryptocurrency trading the following aspects are highly important. First of all, the very nature of cryptocurrency (block chain technology), its novelties and uncertainties make the financial environment surrounding it highly volatile (Chen et al. 2). Figure 1 demonstrates the price fluctuations of Ether currency during the period of 2016 -2017 while Figure 2 further shows similar movement. As prices of cryptocurrencies often and significantly change, the learning algorithms have to be able to account for, prioritize and accommodate multiple variables (Madan et al. 3).

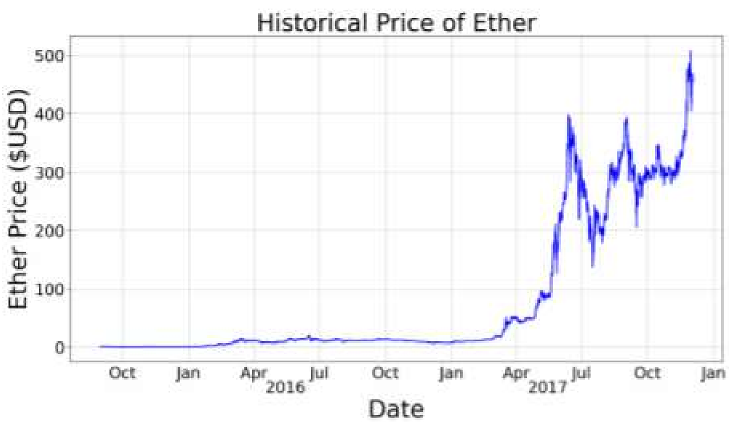

Fig.1. Price changes of Ether (source: Chen et al. 2)

Fig.2. Price changes of Bitcoin (source: blockchain,info)

Tree Ensembles is a family of learning algorithms that have been recently applied to predict crypto pricing, volatility and develop effective cryptocurrency trading models and strategies. The family combines two distinct techniques that will be discussed within the current report: Random Forests and Gradient Boosted Trees (Velankar et al. 3). Random is a model used for machine learning. Random forest can be applied for learning regression, classification and other activities. As indicated by its name, random forests are generated when a large number of decision trees are produced within a specified (training) period of time. Importantly, random forest as an algorithm can autocorrect for tendencies of decisions trees to overfit a specific database (training) set (Velankar et al. 2). Random forest has been discussed as a modification of bagging or bootstrap aggregation, which allows to construct a significant number of various de- correlated trees, while the next step concerns their averaging (Chen et al. 3). The idea behind such approach is to average various highly noisy models and reduce an overall variance of pricing prediction.

Among some other important advantages of random forests as a machine learning algorithm are: it is considered one of the most accurate models available, and good trees are chosen by using greedy algorithms. According to Maidan et al. (5), random forest model can produce a highly accurate classifier. Indeed when compared to the General Linear Model (GLM) algorithm, random decision forest model delivers much more accurate predictive results in relation to cryptocurrency trading. For example, when sampling Bitcoin price changes every ten minutes, Maidan et al. (5) were able to predict the direction in which the cryptocurrency price would change with 57.4% accuracy. However, it is important to point out that although the obtained results are stronger when compared to those of GLM, the obtained prediction rate can still be improved further. Another important issue is that the approach such as random forest allows for almost ready- to – use solutions that can therefore make cryptocurrency prediction available to a wider spectrum of stakeholders (investors) that may not necessarily be able to tailor each parameter (Solanas et al. 35).

However, the approach to machine learning such as random decision forest is also associated with some pitfalls and limitations. Such approach, similarly to gradient boosted trees are highly demanding in terms of computational power and effort (Velankar et al. 36). However, it is important to point out that the two methods differ in terms of CPU needed to compute the task. Random forest method creates decision trees independently and uses only two parameters (number of trees and number of features). On the contrary, the process of tree construction in gradient boosted tree is more complex as it involves using new trees to correct mistakes made by those trees that were trained previously. Thus with addition of every new tree, the process becomes more complex and effort consuming. Therefore machine training based on gradient boosted tree mechanisms is generally significantly longer when compared to random forest one (Solanas et al. 29).

In addition to that, since the random forest model attempts to reduce initial noise, the model is highly sensitive to the quality of the input information. Any biased or incomplete data concerning market volatility, pricing dynamics, etc. may therefore lead to inappropriate models (Madan et al. 4). Such dependence on the quality of input can therefore lead to significant variance and inconsistencies regarding model performance. For example, as shown within Figure 2, when Chen et al. (3) tested different methods (e.g., logistic regression, naïve byes, support vector machines, ARIMA, neural networks) the obtained results indicate that random forest could offer the lowest level of accuracy.

Method	Accuracy
Logistic Regression	53.40 %
Logistic Regression [Binary]	56.94 %
Naive Bayes	51.78 %
Support Vector Machines	51.29 %
Support Vector Machines [Change]	52.59 %
Support Vector Machines [Binary]	55.99 %
Random Forest	50.81 %
ARIMA	61.17 %
Recurrent Neural Network	52.43 %
Neural Network	52.18 %

Fig. 2. Results of using different learning algorithms to predict Ether (source: Chen et al. 2)

According to Chen et al. (5), in relation to cryptocurrency trading and price prediction, random forests and gradient boosted decision trees are distinctly different. Gradient boosted trees, as demonstrated by empirical evidence, allow to maximize model performance. At the same time, such models are significantly more difficult to fine tune as they require input of multiple parameters. The latter can be regarded as a significant pitfall of the method in relation to cryptocurrency pricing prediction. As discussed above, while random forests help avoid model overfitting, gradient boosted decision trees, on the contrary, are highly prone to such problem. The discussed issue, however, can be addressed via applying some technical approaches. For example, Solanas et al. advise to use two parameters such as depth tree and learning rate (or shrinkage) to account for possible model overfitting. Within the context of cryptocurrency price prediction, model overfitting is a serious limitation because it may create a situation where model responds to various types of data noise.

Madan et al. (6) highlight that apart from choosing a specific algorithm, it is pivotal to design the appropriate strategy for the process of trading. For example, it is not sufficient simply to know whether cryptocurrency price will go up or down in the forecasted interval. Instead, it is pivotal to account for the degree of change in the currency price as well as the price that would be the most appropriate to make a buy or sell type transaction (Madan et al. 6). Madan et al. (6) advise to choose for a longer (e.g., 10 minute interval) as opposed to shorter ones, as there is a higher probability to encounter higher price during longer period of trading time.

Conclusions

The current report concludes that the discussed family of algorithms - Tree Ensembles despite some limitations can be regarded as effective approaches to design effective crypto currency trading strategies. Both Random Forests and Gradient Boosted Trees are currently widely adopted to predict pricing of different types of cryptocurrencies, along with other methods such as GLM, Bayesian analysis, neural networks. The analyzed body of literature suggests that random forest can be regarded as one of the most accurate methods, a major benefit of which is that it allows to avoid or reduce model overfitting. The approach however, is highly sensitive to the quality of input data, which is highly important within the context of cryptocurrency trading with its highly volatile and dynamic environment. Another important issue identified within the scope of using different algorithms is model overfitting. The analyzed evidence suggests that random forest can be regarded as a method that is significantly less likely to result in model overfitting when compared to gradient boosted trees. Both methods, however, due to the complexity and excessive number of generated trees lead to significant computational ability needed to predict cryptocurrency pricing based on information concerning cryptocurrency performance, market volatility, etc.

Works cited

Chen, Matthew, Neha Narwal, and Mila Schultz. "Predicting Price Changes in Ethereum."

Madan, Isaac, Shaurya Saluja, and Aojia Zhao. "Automated Bitcoin trading via machine learning algorithms, Department of Computer Science." 2015.

Solanas, Marc, Julio Hernandez-Castro, and Debojyoti Dutta. "Detecting fraudulent activity in a cloud using privacy-friendly data aggregates." *arXiv preprint arXiv:1411.6721* (2014).

Velankar, Siddhi, Sakshi Valecha, and Shreya Maji. "Bitcoin price prediction using machine learning." *Advanced Communication Technology (ICACT), 2018 20th International Conference on*. IEEE, 2018.