Cryptocurrencies Prices Forecasting With Anaconda Tool Using Machine Learning Techniques

Oleksandr Snihovyi[1], Oleksii Ivanov[1] and Vitaliy Kobets[1][0000-0002-4386-4103]

[1] Kherson State University, 27, Universitetska st., Kherson, 73000, Ukraine
snegovoy@hotmail.com, sink2385@gmail.com, vkobets@kse.org.ua

Abstract. The research goal is to study criteria which affect the price of a cryptocurrency and their usage for the price forecasting using Machine Learning techniques. The subject of research is forecasting of most famous cryptocurrency Bitcoin using the most popular Python Data Science Platform – Anaconda. Research Methods are machine learning, data analysis, and multiple regression. A set of criteria which affect the price of the cryptocurrency were defined based on the analysis of public information. Their correctness was tested using Machine Learning algorithms. As a result of simulation experiment through the application using real data from open sources, we have found that selected combination of criterion can explain more than 70% of cryptocurrencies prices variation using either Multiple Regression or Random Trees or Long Short-Term Memory networks.

Keywords: cryptocurrency, bitcoin, forecasting, machine learning, multiple regression, random forests, lstm, long short-term memory, python, data science.

1   Introduction

The story of cryptocurrencies had started in 2009 when Bitcoin was born. It was the first decentralized digital currency with emphasis on cryptography [1]. After nine years, almost 1500 live cryptocurrencies exist [2], and their number continues to overgrow [3].

Another trending sphere that is developing rapidly in the last few years is Machine Learning (ML). It is a method of data analysis that automates analytical model building. ML allows applications to find hidden insights without being explicitly programmed [4]. One of the most common task ML solves correctly is forecasting.

The purpose of the paper is to formalize criteria which affect the price of cryptocurrencies and use them in the price forecasting experiment developed with Python Anaconda Data Science tool with ML algorithms.

The paper is organized as follows: part 2 describes dataset and ML algorithms used in an experiment and the results of the experiment, and the last part concludes.

2 Cryptocurrencies Dataset and ML Prediction Algorithms

Information about cryptocurrencies like their prices per date, their supply, mining difficulty, and other is open source. So we combined data into one dataset [5]. It has next columns:

- date, from 25th of January 2017 to 22 of January 2018 separated by weeks;
- price - the price of Bitcoin for each date (in USD);
- supply - the Bitcoin's total number of coins for each date;
- difficulty - the Bitcoin's mining difficulty for each date (hash rate)[8-hashrate];
- trading_volume - the Bitcoin's trading volume for each date;
- reaction - average society reaction on Bitcoin for each date ("-1" is negative, "1" is positive, and "0" is neutral);

Our dataset has 105 rows, and it was separated into two subsets: training and testing. 70% of data is training set (it is 73 rows) and 30% if data is verification set (it is 32 rows).

In our research we decided to use three supervised ML algorithms to verify if these criteria can be used in cryptocurrency's price prediction in such combination:

- Multiple Linear Regression with default configuration sklearn configuration;
- Random Forests with 100 trees;
- Long Short-Term Memory Networks with 50 neurons on a hidden layer and 100 epochs of training.

Linear regression shows relations between variables and how changes affect them. Random Forests algorithm uses a bagging approach to create a bunch of decision trees with a random subset of data. LSTMs are capable of learning long-term dependencies. The research question is: Which criteria affect the price of the cryptocurrency and how? We have found next points:

1. Total number of mined coins;
2. Mining difficulty level;
3. Cryptocurrency's trading volume;
4. Perceptions of the cryptocurrency's value by the society;
5. Price of Bitcoin.

After all three predictions, we were able to find mean squad error and coefficient of determination. Linear regression application output was:

```
Mean squared error: 2274869.02
R^2 score: 0.79
```

But, Random Forests application output was:

```
Mean squared error: 585482.78
R^2 score: 0.97
```

And, finally LSTM network application output was:

```
Mean squared error: 3248292.00
R^2 score: 0.80
```

In the graphic below we display the results of both predictions in contrast to real values. (fig. 3).
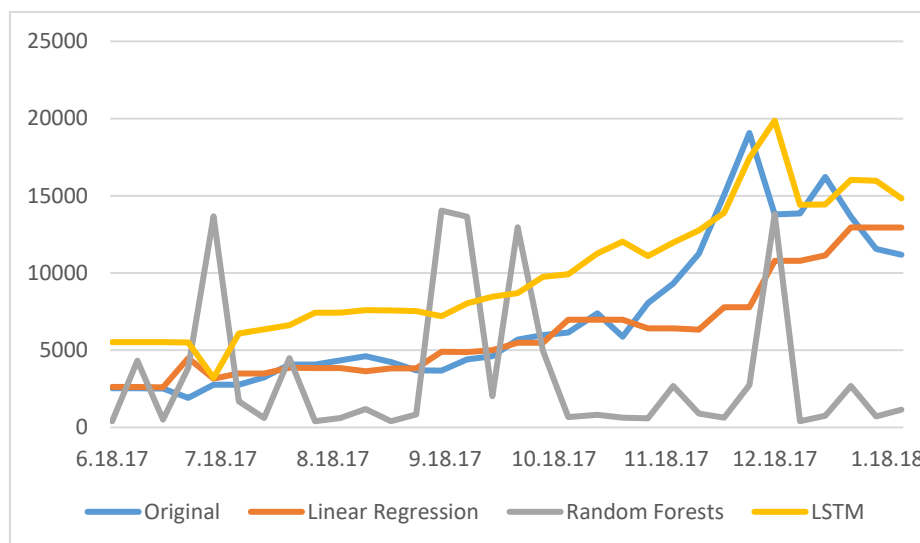

Fig. 1. Comparison of real Bitcoin prices, predicted with Linear Regression and predicted with Random Forests [6]

3   Conclusions and Outlook

Thus we have tested the correctness of the selected criteria combination on their effect on the price of Bitcoin. For our experiment, we used Multiple Linear Regression, Random Forests, and LSTM ML algorithms implemented with Python in Anaconda Data Science tool. As a result, we have found that selected combination of criterion can explain more than 70% of Bitcoin's price.

On this base, we plan to study additional criteria which affect prices of cryptocurrencies to be able to forecast their prices more accurate.

References

1. Bitcoin: A Peer-to-Peer Electronic Cash System, https://bitcoin.org/bitcoin.pdf, last accessed: 2018/1/27.
2. All cryptocurrencies, https://coinmarketcap.com/all/views/all/, last accessed: 2018/1/27

3. 2018 Will See Many More Cryptocurrencies Double In Value, https://www.forbes.com/sites/kenrapoza/2018/01/02/2018-will-see-many-more-cryptocurrencies-double-in-value, last accessed: 2018/1/27
4. Nilson, N.J.: Introduction to Machine Learning. An early draft of a proposed textbook, Robotics Laboratory Department of Computer Science, Stanford University, Stanford, CA 94305 (2005)
5. Dataset, https://drive.google.com/open?id=1pBCtHw4DEGJnqiy-R8Ko3cgD2NRS-qw8
6. Chart dataset, https://drive.google.com/open?id=15CbWUZ2d-ZLYoYentP5qGcF4GWgdn8yG