Estimation of future bitcoin value using historical data and social media sentiment analysis

GROUP 16

DILAN GECMEN - 4221168, JASPER HEMMES - 4226216, LUUK VAN LITSENBURG - 4282345,

JARI LUBBERDING - 4291867, OMBRETTA STRAFFORELLO - 4706889, FREDERIEK WESEL - 4581202

April 24, 2018

I. INTRODUCTION

In the last year, the relevance of cryptocurrencies has been on the rise, and after the value of bitcoin exceeded 9000$ many people started asking themselves if the trend would continue. Given the fact that it is a lucrative and risky investment, we will be investigating whether it is possible to predict the future value in USD of bitcoins by using bitcoin historical data and performing sentiment analysis on social media platforms by making use of a to-be-decided neural network. The influence exerted by social media platforms on certain topic is controversial, take as example the last U.S. presidential election, in which the many social media influencers aligned themselves against a certain candidate, who actually managed to win the election. By doing so they contributed to skew the electoral polls. It will thus be interesting to see if neural networks will be able to capture links between emotion and value of an asset [9].

II. RESEARCH QUESTION

In this project we would like to perform using a neural network of the future (in a to-be-defined time frame) value in USD of bitcoin, based on historical data and sentiment analysis of social media content related to the subject, and see if there is a significant gain in performance compared to the estimation performed with historical data exclusively. Emphasis is placed both on the sentiment analysis as well as the estimation.

III. EVALUATION

Since our results consist of predictions, they will be the data-points of the value of bitcoin in US Dollars in the future. If we work using past data the algorithm will predict a value which we can simply compare to the actual value of bitcoin that day. To evaluate how well the predictions will be compared to ground truth, an error will be defined, most likely as the p-norm of the difference between prediction and ground truth. Based on this, error metrics such as

spread can be derived. Furthermore, standard comparisons between training, cross-validation and test error will be computed, in order to understand if we are in presence of low or high bias/variance. Of course these estimates will be performed multiple times, in order to guarantee statistical relevance to our results. Creating estimations using only historical bitcoin data and estimations using historical bitcoin data and the sentiment analysis we can answer the research question.

IV. METHOD & DATA

The first model will be a model estimating future values of bitcoin using historical bitcoin data. The dataset by crypto datasets will be used for the bitcoin price [1], also transaction data is available [6]. This collected data will be used in a neural network to predict future values of bitcoin. This data will be processed to extract features which serve as input for the network. Bayesian regression together with neural networks have proven to be successful for bitcoin price prediction in the past [4] [10] and therefore will be considered. An alternative method that will be evaluated for use are Recurrent Neural Networks composed of Long Short-Term Memory (LSTM) units, as they have proven to be successful as well for this purpose [8]. Besides Bayes and RNN composed of LSTM units, Convolutional Neural Networks (CNN) are also considered as they have proven to be useful in portfolio management of cryptocurrency [5].

The method that will be chosen from the paragraph above will be extended using a sentiment analysis of social media posts or Google trend statistics. Either by processing the data into an average value for sentiment of the public and feeding that value as a feature to the network, or by scaling the estimated output of the network dependent of the sentiment of the public. The data used for the sentiment analysis is publicly accessible online data that will be collected from; Twitter, Google Trends, Facebook, and Instagram. This data will contain messages relevant to the subject, for example by using all posts which contain #bitcoin, also data about the frequency of people searching for bitcoin can

be used. The data from twitter will be collected through Tweepy [11], which is a python library that helps linking Python scripts with Twitters API. The data from Google trends will be collected through PyTrends [7], which is a python library that links python scripts with Google Trends. The data from Facebook will be extracted using the official Facebook Graph API [3]. Finally, for Instagram the data will be obtained using an unofficial python API [2].

V. Timeline

The work is going to be divided in three main tasks:

1. development of the sentiment analysis system, based on data from social media posts or Google trend statistics;

2. construction of the neural network used to predict bitcoin trends from historical data;

3. integration of the two implementations and evaluations of the results.

Each part is going to be assigned to two of us, in such a way to split the workload evenly, and respect the following timetable.

2018-04-23	Start collecting data
2018-04-30	Sentiment analysis results
2018-05-07	Working implementation of the RNN
2018-05-14	Intermediate results
2018-05-21	Milestone report
2018-06-04	Final implementation and results
2018-06-11	Project completion
2018-06-18	Work on the poster

VI. Feasibility

Overall the project should be feasible. In particular there are many, easy-to-access databases containing historical data of cryptocurrencies. It will be a bit more challenging to come up with historical sentiment analysis, but we think we should be able to address the problem, given that is is becoming increasingly popular. Combining the sentiment analysis with the value estimation using historical bitcoin data is probably the most challenging as we did not find any article doing so, however we feel the proposed methods can work. Designing, testing a suitable neural network is also something that should be relatively easy to accomplish, as all of our team has moderate programming skills. Although there is only one student with a Computer Science background, all team members have finished the lab exercises in a decent amount of time and therefore no problems are expected based on this evaluation criterion.

VII. Future work

If successful it might be interesting to extend this approach to multiple cryptocurrencies and create an algorithm that automatically manages a portfolio of currencies, which might be able to give a very high return on investment.

References

[1] CryptoDatasets. Bitcoin price dataset, 2018. Available at `https://www.cryptodatasets.com/`.

[2] Tom Dickinson. Instagram search api for python, 2018. Available at `https://github.com/tomkdickinson/Instagram-Search-API-Python`.

[3] FaceBook. Graph api, 2018. Available at `https://developers.facebook.com/docs/graph-api`.

[4] H. Jang and J. Lee. An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information. *IEEE Access*, 6:5427–5437, 2018.

[5] Z. Jiang and J. Liang. Cryptocurrency portfolio management with deep reinforcement learning. In *2017 Intelligent Systems Conference (IntelliSys)*, pages 905–913, Sept 2017.

[6] MIT. Bitcoin network dataset, 2018. Available at `https://senseable2015-6.mit.edu/bitcoin/`.

[7] PyTrends. Unofficial api for google trends, 2018. Available at `https://github.com/GeneralMills/pytrends`.

[8] D. G. Torres, H. Qiu. Applying recurrent neural networks for multivariate time series forecasting of volatile financial data. Technical report, KTH Royal University of Technology, 01 2018.

[9] S. Colianni, S. C. Rosales and M. Signorotti. Algorithmic trading of cryptocurrency based on twitter sentient analysis. 2015.

[10] D. Shah and K. Zhang. Bayesian regression and bitcoin. In *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 409–414, Sept 2014.

[11] Tweepy. Tweepy documentation, 2018. Available at `http://tweepy.readthedocs.io/en/v3.5.0/`.