Capstone Project Proposal
Machine Learning Engineer Nanodegree

"Using Machine Learning Models to Forecast Bitcoin Price Movements"

Author: René Javier Rivero Arrieta
Essen, Germany
03.11.2020

Domain Background

Since the 20th century, mathematicians, economists, business analysts and traders have used computers to come up with different technical and fundamental indicators to predict the price movements on the stock market. In one article, The Economist ('The Rise of the Financial Machines', 2019) mentioned that algorithmic funds "account for 35% of America's stock market, 60% of institutional equity assets and 60% of trading activity", which is an impressive number for the amount of money that means.

More recent academic research has also explored the use of different Machine Learning algorithms to predict stock price movements. While one may find a lot of academic sources, I would mention some recent studies like 'Prediction of Stock Market by Principal Component Analysis' (Waqar, Muhammad & Dawood, Hassan & Guo, Ping & Shahnawaz, Muhammad & Ghazanfar, Mustansar ali., 2017), 'Forecasting daily stock market return using dimensionality reduction' (Zhong X. Enke D, 2017), 'High-performance stock index trading via neural networks and trees' (Chalvatzis, C., Hristu-Varsakelis, D., 2020) or 'An innovative neural network approach for stock market prediction' (Pang, X., Zhou, Y., Wang, P., Lin, W., Chang, V., 2020), all exploring different algorithms and approaches to stock market prediction.

Problem Statement

Financial products have evolved, as algorithms and technology also have. With relatively short span of life, one the best performing financial asset is considered to be Bitcoin, created in 2008 by an author (or group of authors) with the pseudonym Satoshi Nakamoto ('Bitcoin: A Peer-to-Peer Electronic Cash System', 2008). However, due to the its short and also accelerated growth, Bitcoin has not been studied as other more mature financial assets. For this reason, my objective with this Capstone Project, will be to leverage on different machine learning algorithms to, as a first step, understand the principal components affecting the movements of Bitcoin prices, and as a second step, create a price forecaster or predictor, that I will use for predicting Bitcoin price movements.

Datasets and Inputs

The datasets to be used in this study will mainly rely on stock market price data and fundamental Bitcoin data publicly available in different sources, such as Quandl, TradingView, CoinMarketCap and Investing.com. Some of the characteristics of all these data sets are:

- They are all time series data
- The timeframe taken for the data points will be daily
- The start date will be since the beginning of Bitcoin in 2008 (wherever data is available)
- Price data usually in OHLC (open, high, low, close) feature format. For our purposes we may concentrate on the 'closing' values if others are not considered relevant. Daily closing values will also be used for further calculations needed.

Some of the features considered in the input data sets will be:

- Stock market: Bitcoin price and volume, SPX500, NASDAQ, Dow Jones, DAX30, gold price, silver price, oil price, interest rates (wherever available)
- Bitcoin fundamental data: hash-rate, block size, transaction fees (wherever available)

Additional technical indicators, such as moving averages and relative strength index (RSI), may also be included in the model but will need to be calculated during the project.

Solution Statement & Benchmark Models

From the set of daily price and technical indicators, we will use Principal Component Analysis to determine the most relevant data vectors affecting Bitcoin price and select appropriate data inputs for our price estimator at a later stage.

Once the principal components affecting Bitcoin price have been identified using PCA and feature engineering, I will use them to train a model to either forecast, or estimate 'buy', 'sell' or 'no trade' signals optimizing intra-day returns as the optimization variable. For this, my benchmark model will be XGBoost, given its good and lightweight performance, and additionally I will be evaluating a Neural Network against it, since consulted bibliography have proven good results for price predictions of other assets in the past.

Evaluation Metrics

In order to evaluate my model(s), I will compare the intraday-returns of allocating equity to this predictor over time, in comparison to allocating equity to Bitcoin without using the model. This will first allow us to evaluate if the returns of using the model are actually due to the model performance or if the returns observed are simply due to the growth in Bitcoin prices. This analysis will be done in different market scenarios, i.e. a bull market (period of price increase), a bear market (period of price decrease) and sideways behavior.

When comparing the two models (XGBoost and Neural Networks), the one that performs better will have better cumulative returns over time than the other, being this my main evaluation metric. I will also observe predicted prices on the next day for a number of days and evaluate their accuracy towards the predicted price movement.

Project Design

While the approach adopted will be the one mentioned before, it may change 'on-the-go' if hard blockers are found or if better solutions or approaches are found during the execution. More concretely, the following steps will constitute the project workflow and any changes will be explained in the relevant project notebook(s):

- Data sourcing, cleaning and normalization
- Calculation of additional technical indicators
- Principal Component Analysis and Feature Engineering
- Bitcoin price forecasters and predictors: XGBoost vs Neural Networks
- Model evaluation and results discussion

By the end of the project we would have a better idea of the different variables affecting the price of this revolutionary asset, and a predictor of its price movements. The rights for publishing and use of the resulting model(s) and result(s) will be reserved by the author.

References

Chalvatzis, C., Hristu-Varsakelis, D., 2020. *High-performance stock index trading via neural networks and trees.* Applied Soft Computing, Vol. 96, November 2020, 106567.

Pang, X., Zhou, Y., Wang, P., Lin, W., Chang, V., 2020. *An innovative neural network approach for stock market prediction.* The Journal of Supercomputing 76, 2098-2118(2020).

Satoshi Nakamoto, 2008. *Bitcoin: A Peer-to-Peer Electronic Cash System*. https://bitcoin.org/bitcoin.pdf

The Economist (2019) *The Rise of the Financial Machines.* The Economist Group Limited, London 2019. https://www.economist.com/leaders/2019/10/03/the-rise-of-the-financial-machines

Waqar, Muhammad & Dawood, Hassan & Guo, Ping & Shahnawaz, Muhammad & Ghazanfar, Mustansar ali., 2017. *Prediction of Stock Market by Principal Component Analysis*. 13th International Conference on Computational Intelligence and Security (CIS).

Zhong X. Enke D, 2017. *Forecasting daily stock market return using dimensionality reduction.* Expert Systems with Applications, Vol. 67, January 2017, Pages 126-139.