
Analysis and Prediction of Cryptocurrency Prices

Péter Garamvölgyi, 2018280070¹ Federico Zaiter, 2018280071¹

Abstract

Stock market price prediction is a well-researched problem with numerous applications of machine learning models. However, research on how to apply such models on cryptocurrency prices is still in its early stages. In this project we evaluate some common time-series analysis models with the goal of predicting Bitcoin prices.

1. Problem statement

Statistical analysis has been applied to predict stock pricing for a long time. In the recent decades, various machine learning models have been successfully applied to such time-series analysis problems. Commonly applied methods include various regression models, RNNs (Recurrent Neural Networks), and even Reinforcement Learning.

Inspired by the Kaggle challenge *Bitcoin Price - Build time series models to predict bitcoin price*¹, we set out to explore the applicability of these methods for predicting cryptocurrency pricing. While similar at first sight, the price profiles of these assets demonstrate some unique features compared to the classic problem of stocks: non-stop trading, high volatility, stronger susceptibility to the general sentiment among others.

Our main goal is to develop a model that is able to predict the price of Bitcoin based on historical data with reasonable accuracy. A follow-up question is: Based on this model, is it possible to create a profitable trading policy?

2. Problem model

A simple model of predictive time-series analysis is as follows. We are given a set of training data consisting of observation vectors and price labels:

¹Department of Computer Science and Technology, Tsinghua University, China. Correspondence to: Péter Garamvölgyi <garamvoelgyip10@mails.tsinghua.edu.cn>, Federico Zaiter <zaitertf10@mails.tsinghua.edu.cn>.

¹<https://www.kaggle.com/c/bitcoin-price>

$$\{x_i, y_i\}_{i=1}^N \quad x_i \in R^D, y_i \in R$$

Our goal is to train a model on this dataset so that our model has reasonable generalization abilities on future or otherwise unlabelled data points.

3. Our approach

Our approach consists of 5 main steps:

1. Data acquisition
2. Data exploration
3. Preprocessing and feature extraction
4. Model fitting and training
5. Evaluation and comparison

For data acquisition, we will first rely on the example Bitcoin dataset provided on Kaggle. In the future, we might consider extending our solution by using more recent data or data from other cryptocurrencies.

For data exploration we will use Pandas, matplotlib, and other Python tools. Similarly, we will rely on well-established Python libraries like tsfresh for feature extraction.

The heart of any machine learning application is the actual model used. We are planning to evaluate the ARIMA (Auto-Regressive Integrated Moving Average) statistical regression model, LSTM (Long-Short Term Memory Units), and the tools provided on Azure Machine Learning.

Finally, we will provide a comparison of the applied models.

4. Expected results

Our baseline is the naïve forecasting approach: In every prediction step, we simply use the previous label as our prediction without any adjustment.

Our goal is to implement a model with superior performance compared to the baseline. The performance can be evaluated using various metrics like RSME or MAE.