University of St.Gallen

School of Management, Economics, Law, Social Schiences and International Affairs

The effect of Twitter activity on Bitcoin price

Documentation

Dimitrios Koumnakes - 10-613-370
Severin Kranz - 13-606-355
Joël Sonderegger - 11-495-488
Alen Stepic - 11-475-258
Chi Xu - XX-XXX-XXX

Fall Term 2017

Supervisor
Prof. Dr. Philipp ZAHN
Department of Economics

December 27, 2017

Abstract

(Insert text)

Contents

List of Figures

1 Introduction

In academics accountability and reproducibility is important. However, the publishing process of papers and journals seem to be outdated, as many new ways of data collection and processing exist. Examples of academic papers, which are often cited exist where the initial computation are not reproducible and contain errors. Even though just replicating data or existing results do not provide any new knowledge at all, it gives a good indication about the quality of the work done and hence increases trustworthiness. This explains why reproduction is of great relevance and point out the existing problems.

1.1 Goal of the paper

The goal of this documentation paper is to provide a description how the data discussed in the separate paper have been gathered, stored, aggregated and analysed. By reading the documentation the reader should be able to reproduce the results.

1.2 Methodology

This documentation consists out of four chapters. The first chapters contains a short introduction and provides the reader with an overview about the topic. Furthermore it points out the relevance of documentation. The second chapter contains an description how the twitter tweets were collected and stored. The third chapter discusses how the bitcoin data is gathered. Finally the fifth chapter discusses how the analysis has been conducted.

1.3 Scope

The scope of the documentation is provide an overview about the different steps which have been conducted to obtain the results in the paper. It does not contain any discussions about the results of the separate academic paper. It is not a deep description of the code as the code itself was documented separately. Nevertheless, important lines of code are discussed.

2   Data Collection

Here, we provide a detailed description of how the data for the sequential analysis is gathered and stored. This includes two subsections the (1) tweets data and the (2) bitcoin price data.

2.1   Tweets Data

To collect the needed twitter data a python script has been written. Whit the python script real-time twitter data are streamed and stored by the usage of a raspberry pi.

2.1.1   Python Script

Twitter offers different Application Programming Interfaces (API) for collecting data. Based on twitter, the biggest an probably most relevant API's are (1) Ads API, (2) Filter realtime Tweets, (3) Search Tweets and (3)Direct Message API. (How to add a reference???) However, the timeframe for gathering data for free is limited to 7 day's. As python offers different libaries tweepy a open-source libary is been used for streaming the data. By using tweepy twitter can be reached in two way's (??) or by the streaming API. The latter has been used in contex of the assignment.

Stream - push then pull, easier - authentic

Twitter Authentification   - printscreens and how to proceed - pip install??

Twitter Streaming API   ....

2.1.2   Hardware Setup

(Severin's Part)

2.2   Bitcoin Price Data

We wrote a Python script which collects Bitcoin price data as there was no preexisting data collection that satisfied our needs. The Bitcoin price is best expressed by the Bitcoin Price Index. The Bitcoin price index (BPI) is an index of the exchange rate between the Bitcoin (BTC) and the US dollar (USD) (**kristoufek2015main**). The objective of the script was to gather hourly Bitcoin Price Index data for at least the time period in which we gather the tweets data. We found the an API by bitcoinaverage.com which sufficed our needs. An API description follows later.

2.2.1 Execution

By executing the python script `collectCryptocurrencyData.py` hourly data for the Bitcoin Price Index is retrieved.

```
$ python collectCryptocurrencyData.py
```

2.2.2 Output

After successfully running the python script `CollectCryptocurrencyData.py` the file `bpi.csv` is generated in the folder `/data`. It is important to note that every execution of the script overwrites any existing `bpi.csv` file.

The file `bpi.csv` contains historical Bitcoin Price Index data for one month on an hourly basis. Each data point consists of the following parameters:

- time: Timestamp on an hourly basis

- average: Average price (in USD)

- high: Highest price (in USD)

- low: Lowest price (in USD)

- open: Opening price (in USD)

2.2.3 API: Bitcoinaverage.com

Bitcoinaverage.com offers a free API that provides real-time and historical price data for a range of crypto-currencies including Bitcoin. The following requests delivers data for an per hour monthly sliding window.

Request
The request to get the data for an per hour monthly sliding window looks as follows. This request require authentication that requires registration and the generation of an API key. The registration and generation of an API key is freely available on bitcoinaverage.com. The collectCryptocurrencyData.py already contains the necessary keys. This means that you need no register or generate keys to execute the script collectCryptocurrencyData.py.

```
https://apiv2.bitcoinaverage.com/indices/global/history/BTCUSD?period=monthly&?format=json
```

Response   An excerpt of an example response looks like the following:

```
[
    {
        "high": 8271.04,
        "average": 8247.83,
        "open": 8242.39,
        "low": 8217.72,
        "time": '2017-11-22 15:00:00'
    },
    {
        "high': 8246.82,
        "average": 8203.19,
        'open': 8203.81,
        'low': 8157.25,
        'time': '2017-11-22 14:00:00'
    },
    {
        'high': 8267.27,
        "average": 8238.62,
        'open': 8248.77,
        'low': 8198.54,
        'time': '2017-11-22 13:00:00'
    }
]
```

The time timestamps is UTC time.

3   Data Aggregation

(Joel's Part)

4   Data Analysis

(Dimitri's Part)