

Report D15

Project title: KAGGLE-Crypto prices

Team members: Erik Presnov, Maria Anett Kaha

GitHub repository link: <https://github.com/ErikPresnov/IDS2022-Crypto>

Dataset: <https://www.kaggle.com/datasets/prasoonkottarathil/btcinUSD?select=BTC-Daily.csv>

Business understanding

Business goals

Accurate stock price prediction is extremely challenging because of multiple (macro and micro) factors, such as politics, global economic conditions, unexpected events, a company's financial performance, and so on. If we find in our results that our project can predict the exact prices of cryptocurrencies, we would make a lot of money.

The aim of the project is to know if we have the ability to predict the future prices of cryptocurrencies. In order to achieve that goal, we compare our predicted prices with the actual prices of Bitcoin on 02.03.2022.

Situation assessment

For this project we will use Jupyter Notebook as the environment. As for computing power, we have three laptops and one desktop computer which will hopefully be enough. The data we are planning to use is a Kaggle dataset.

Requirements, assumptions, and constraints

The only constraint is the deadline for the project. Of course, there is the possibility that what we are trying to do is impossible or not within our capabilities but that is what we are going to find out.

Terminology:

- Bitcoin (BTC)
- All time high (ATH)
- Moving average (MA)
- Relative strength index (RSI)

Costs and benefits

Since there are no real costs to this project except our time and some electricity bill there could be a lot of benefits if this project turns out to be a success. If we are able to predict the direction of the price movement there is already potential to make money. If we are able to predict the exact price (highly unlikely but still a chance) there is even more potential for making money. The issue is whether we are confident enough to trust a computer program to trade with our money.

Data mining

The goal of our data mining is to get data about cryptocurrency. We chose to go with Bitcoin. We looked through some available Kaggle datasets about cryptocurrencies and found one we were happy with.

Data understanding

Data gathering

Kaggle is full of datasets about cryptocurrencies, especially Bitcoin since it is the oldest and most popular one. We chose a dataset which contains hourly (May 2018 to March 2022) and daily (November 2014 to March 2022) data about BTC. We also have access to minutely data from January 2017 - March 2022. Since we have data with three time intervals we could try to make two models: one which predicts the long term movement (a few days) and another one which predicts short term (minutes or hours).

The data contains information about time (either a date or a date and timestamp), opening price of the timeframe, high price of the timeframe, low price of the timeframe, closing price of the timeframe and volume in BTC and USD.

Exploring data

The data we have contains 2 (2017 (19100\$) and 2021 (64150\$)) out of the 3 big all time highs so we have data about most of the ups and downs. The one we do not have data about is the first crypto boom from 2013 (1150\$). Since the price of BTC is now closer to the 2017 boom price we think not having the first ATH data will not be too big of a problem.

Data quality

The data seems to be complete as in no values are missing. Since the data does not contain any indicators we might have to calculate them ourselves.

The ones we might use are:

- Moving average -> The average price over X timeframes, eg. MA200 would show the average price over 200 timeframes.
- Relative strength index -> Indicates whether the stock (BTC in our case) is oversold or overbought. The most commonly seen boundaries are 30/70 meaning if RSI is below 30 the stock is oversold and overbought if RSI is above 70.
- Bollinger bands -> Is a band where the upper boundary is moving average + someConstant*standardDeviation and the lower boundary is moving average - someConstant*standardDeviation. So if the price is touching the band that means the price is "extreme" with respect to the average.

Project plan

Task	Methods/tools	Comments	Time
Choose topic	<ul style="list-style-type: none">- Choose a topic;- Choose dataset;- Make goals	First steps of the project	Both: 2h
Report	<ul style="list-style-type: none">- GitHub repository;- Business understanding;- Data understanding;- Project plan	PDF report	Both: 4h
Data preparation	<ul style="list-style-type: none">- Data cleaning;- Data preparation;- Data transformation	Calculating the above mentioned indicators.	Both: 7h
Data analysis	<ul style="list-style-type: none">- Data analysis;- Data modelling;- Data classification;- Data forecasting	Jupyter Notebook	Both: 7h
Compare the results	<ul style="list-style-type: none">- Compare results to the actual future data(02.03.2022 data)	Goal 1: try predicting the prices of cryptocurrencies in the future. Goal 2: try predicting the direction, where the price moves.	Both: 3h
Reporting final results	Make poster PDF of the project	Final report and presentation	Both: 7h