**Data Analytics and Data Driven Decision**

Academic year 2017-2018

**University of L’Aquila, Italy**

**Cryptocurrencies**

Ebrima M. Bah - 254906

Fabio Giancristofaro - 255146

Hilmi nurArdian - 255449

Stefan Ilic - 252212

**Table of contents**

1. ***Introduction*** *3*
   1. ***1 What is a cryptocurrency?*** *3*
   2. ***Why this data?*** *3*
2. ***Description of dataset*** *4*
3. ***Exploratory analysis*** *5*
4. ***Unuspevised learning*** *6*

***4.1 Metcalfe’s law***

1. ***Supervised learning*** 7
2. ***Conclusions 8***
   1. **Introduction**
   2. **What is a cryptocurrency?**

Cryptocurrency is a kind of [digital currency](https://en.wikipedia.org/wiki/Digital_currency), [virtual currency](https://en.wikipedia.org/wiki/Virtual_currency) or [alternative currency](https://en.wikipedia.org/wiki/Alternative_currency). Cryptocurrencies use decentralized controlas opposed to centralized [electronic money](https://en.wikipedia.org/wiki/Electronic_money) and [central banking](https://en.wikipedia.org/wiki/Central_bank) systems. The decentralized control of each cryptocurrency works through [distributed ledger](https://en.wikipedia.org/wiki/Distributed_ledger) technology, typically a [blockchain](https://en.wikipedia.org/wiki/Blockchain), that serves as a public financial transaction database.

[Bitcoin](https://en.wikipedia.org/wiki/Bitcoin), first released as open-source software in 2009, is generally considered the first [decentralized](https://en.wikipedia.org/wiki/Decentralization) cryptocurrency. Since then, over 4,000 *altcoin* (*alternative coin*) variants of bitcoin have been created.

Cryptocurrencies are used primarily outside existing banking and governmental institutions and are exchanged over the Internet. While these alternative, decentralized modes of exchange are in the early stages of development, they have the unique potential to challenge existing systems of currency and payments. As of April 2018, total market capitalization of cryptocurrencies passes 400 billion USD.

The validity of each cryptocurrency's coins is provided by a [blockchain](https://en.wikipedia.org/wiki/Blockchain). A blockchain is a continuously growing list of [records](https://en.wikipedia.org/wiki/Record_(computer_science)), called *blocks*, which are linked and secured using [cryptography](https://en.wikipedia.org/wiki/Cryptography). Each block typically contains a [hash](https://en.wikipedia.org/wiki/Cryptographic_hash_function) pointer as a link to a previous block, a [timestamp](https://en.wikipedia.org/wiki/Trusted_timestamping) and transaction data.

* 1. **Why this data?**

We choose this dataset because is one of the most discussed topic in the last period. Lot of people is trying to study the trade of these coins, others are mining them. In the month of December 2017, the price of Bitcoin reached 20000 USD, an incredible value especially if we consider that 2 years before the price was near 300 USD.

Cryptocurrency is beginning to indirectly impact traditional exchanges, too. Interest in cryptocurrency and the Blockchain is high enough that any mention of these technologies can [cause a surge in share price](http://www.telegraph.co.uk/technology/2018/01/22/telecoms-minnows-shares-rise-130pc-adding-blockchain-name/).

The aim of this project is to show the trend of cryptocurrencies in the last 3 years, trying to answer to the most asked question: how the value of a cryptocurrency changes?

The most significant information is the value of Bitcoin, the father of all the other cryptocurrencies. We will use Bitcoin as reference for all other cryptocurrencies.

# Actually, there aren’t relevant economic data directly connected with the cryptocurrencies trades but we will try to use different way in order to show some good results.

* 1. **Description of dataset**

Our dataset is composed by different cryptocurrencies information. That’s why it’s probably a good idea to pull the data from one source. We’ll use [coinmarketcap.com](https://coinmarketcap.com).

Each coin has the same 7 columns:

* date
* txVolume
* txCount
* marketcap
* **exchangevolume**
* generatedcoins
* fees

First, we have the **date**. We use the gregorian calendar with the format dd/mm/yyyy. Daily closes for our price quotes occur at 00:00 UTC.

Next, we have **txVolume(USD).**That’s what we’re talking about when we say “on-chain transaction volume.” Simply put, it’s a broad and largely unadjusted measure of the total value of outputson the blockchain, on a given day. This is an answer to the question “approximately how much value, denominated in USD, circulates on the Bitcoin blockchain a day?”.

The third column is **txCount.**That refers to the number of transactions happening on the public blockchain a day. For low-fee blockchains, it’s really easy to fabricate a whole bunch of transactions. Additionally, UTXO networks like Bitcoin can batch a whole bunch of transactions into one, so **txCount**underestimates those ones.

Next, **marketcap(USD).**This is of course the unit price multiplied by the number of units in circulation. Marketcap or network value is definitely flawed. It becomes less tethered to reality the smaller the float is. Float means the ratio of actual circulating units to the total number of units. Ripple, for instance, has a fairly small float, so one should probably be skeptical of its “market cap.”

**Price** is the opening price. We get it from CoinMarketCap.

**exchangevolume(USD)**is the dollar value of the volume at exchanges like GDAX and Bitfinex. It doesn’t include data on OTC exchanges, which is a meaningful portion of all global exchange.

Next, **generatedCoins.**This refers to the number of new coins that have been brought into existence on that day. We count up the actual number of newly-minted coins, rather than using the stated inflation figures (i.e. for bitcoin you should expect 12.5 per block, every ten minutes, giving you 12.5\*6\*24 = 1800 coins per day).

Lastly, **Fees**. Fees in our data are based on the native currency, not USD. Transaction fees for cryptocurrency depend mainly on the [supply](https://en.wikipedia.org/wiki/Supply_and_demand) of network capacity at the time, versus the [demand](https://en.wikipedia.org/wiki/Supply_and_demand) from the currency holder for a faster transaction. The currency holder can choose a specific transaction fee, while network entities process transactions in order of highest offered fee to lowest.

* 1. **Exploratory analysis**
  2. **Unuspevised learning**
  3. **Metcalfe’s law**

Metcalfe's Law states that the value of a network is proportional to the square of the number of users on the network. The classic example is a fax machine: a fax machine is useless by itself but is very useful if a few of your friends have one. If the number of fax machine user’s doubles, the value of the network increases exponentially.

[Fundstrat](http://www.businessinsider.com/bitcoin-price-movement-explained-by-one-equation-fundstrat-tom-lee-metcalf-law-network-effect-2017-10#ampshare=http://www.businessinsider.com/bitcoin-price-movement-explained-by-one-equation-fundstrat-tom-lee-metcalf-law-network-effect-2017-10) had the idea to try and apply this formula to the price of bitcoin. Unique bitcoin addresses are used as a proxy for number of network users. He also added to this formula the number of transactions per user. Fundstrat found a formula by regressing the price of bitcoin against both unique addresses squared and transaction volume per user. This model explained 94% of the variation in the cryptocurrency price since 2013.

*BTC* = *x \* n* 2 *\***t/n*

Where X is a constant, n is number of addresses, and t is transactions

In this notebook, we will try to replicate these results, trying to modify the main formula. We will use the transaction volume instead the t/n value.

*BTC* = *x* \* *n* 2 \* *v*

Price prediction is done in different way, splitting the…

And using one and more coins.

* 1. **Supervised learning**
  2. **Conclusions**

We can see that prediction with Metcalfe’s law improve comparing more than one coin.