**Getting Data**

First part of the project is concerned with obtaining data. As the market is still young and volatile, only cryptocurrencies with enough price history and high trading volume are considered. Higher volume represents more liquidity on the market, which is essential as it ensures that the price is not stagnating and at the same time also dampens possible pumps and dumps. Fundamental side of a project behind each cryptocurrency is also considered. I focused on projects that already had their main net launch or will be doing one in the near future. As the majority of projects start on Ethereum blockchain to shape the product and raise capital, they later transition to their own blockchain once it is running stable. This is known as the main net launch and is proof of a working product.

After taking these criteria into account, the search has been narrowed to three coins, namely EOS, TRX and ONT, which all differ in age of the project as well as total market capitalisation. Out of over 1600 cryptocurrencies, the three selected are all in top 25 by market cap and 24-hour volume traded according to Coinmarketcap[[1]](#footnote-1).

**EOS**

EOS is currently sitting on 5th place by market capitalisation, with valuation of approximately 9.30 billion USD and daily traded volume of 900 million USD.

Earlier this month, the project completed their initial coin offering, or ICO, and raised over a billion USD. It was the longest lasting ICO to date that began in July 2017. On June 2nd, 2018 main net was successfully launched and EOS blockchain went live. The project aims to provide a platform for building decentralised applications, or DApps, similar to the Ethereum blockchain with some key improvements:

* It uses Delegated Proof of Stake (DPoS), which is believed to be less centralised and operate more efficiently, compared to Ethereum’s Proof of Work (PoW).
* At launch EOS could process around 1000 transactions per second (tps), with potential to increase the number by horizontal scalability – processing transactions while simultaneously executing smart contracts. Ethereum blockchain on the other hand still averages at 15tps.
* DApps on EOS blockchain can be developed in any language that is compiled in WebAssembly (e.g. C++), compared to Solidity, which is a proprietary programming language that is used to develop applications on the Ethereum blockchain.
* Broken DApps on the Ethereum network require a fork and cannot be modified otherwise, however, EOS’s DPoS allows to freeze and resume the network after a fix has taken place.

**TRX**

TRX is ranked 10th with 2.90 billion USD market capitalisation and 450 million USD volume traded daily.

The project aims to disrupt the global entertainment industry (currently valued at around a trillion USD) by providing a decentralised content sharing platform. It enjoys massive support in China, where it is also based, with some key employees from Alibaba and Tencent and is backed by notable members of Chinese business community. They completed ICO in August 2017 and launched main net on the 31st May 2018.

**ONT**

ONT is the youngest of the three and received its first major exchange listing only a few months ago (listed on Binance in March 2018). Nevertheless, it is already valued at 930 million USD, taking 19th place by market capitalisation. With 150 million USD it also has the lowest 24-hour volume traded out of three selected coins.

The project focuses on removing barriers between the blockchain and the business sector, hoping that even without knowledge of blockchain business can utilise and integrate the underlying blockchain technology into their products. Ontology (ONT) wants to serve as a trust layer and resolve issues such as identity verification, privacy protection – to tackle current lack of ownership of personal data, data management monopolisation, identifying false information and more. The project aims to create an ecosystem for private blockchains to exchange data for verification, reputation and collaboration, whilst serving as a communication protocol between them.

The above cryptocurrencies were selected as they had consistent trading volume over the past two months and show strong fundamental development, which I believe is vital for adoption and future growth. After 70% correction from the highs of total cryptocurrencies market valuation in January 2018, EOS’s price had already surpassed its all-time high. On the other hand, the price of TRX, that saw a 22.5x increase, still sits 65% lower than in January, and OTN was first listed on an exchange after the bull run and correction with its price steadily growing. I believe that combination of price history with fundamental analysis of each project presents three unique standpoints for time series analysis using machine learning.

**Exchanges & APIs**

The nature of cryptocurrencies is to keep data decentralised and publicly available. The historical price data can be therefore obtained directly from major exchanges. At first, I intended to use the Poloniex’s public API as it provides price history in up to 5min intervals with no restrictions on the number of requests or size of data requested. Although they support trading of more than 100 cryptocurrencies, they have yet to list any of the three selected coins.

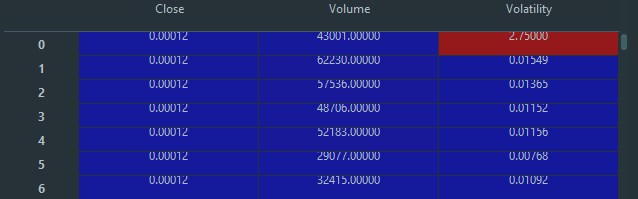
Due to this limitation I decided to use Binance exchange as they also provide a public API, hence there is no requirement to open an account with them just to download the price history data. They also offer more variety in candlestick duration – intervals span from 1 minute up to a month, however, due to the high market volatility this project will focus on smaller time frames. For flexibility I decided to work with 5m, 15m and 30m intervals of historical price data. Contrary to Poloniex, Binance API returns max 500 records per request (e.g. 500 price points) and limits the number of requests to 1200/minute before an IP ban. This figure is still far greater than the available price history and even the existence of Binance exchange, since 1200 of 5min interval requests return around 5.7 years of price data points. Both exchanges return data in JSON format, making it easy to manipulate with Python’s core json library.

As I will be working with LSTM neural network that requires a large number of input points, the complete dataset for each cryptocurrency is obtained – from their listing on Binance exchange until 15th June 2018 at 00:00:00. Their API works with timestamps from epoch, hence all dates were first converted into a proper format. As the API returns max 500 price points per request they were concatenated and saved into a .json file for faster subsequent lookup. Collected datasets are small and of size <25MB, hence requests’ returns were stored in memory and written to a file once, not at each request as IO functions are time-heavy.

**Processing Data**

As part of data cleaning, all rows containing null values or a timestamp greater than the one of 15th June 2018 at 00:00:00 are removed.

The JSON objects returned by Binance consist of 12 fields: Open time, Open, High, Low, Close, Volume, Close time, Quote asset volume, Number of trades, Taker buy base asset volume, Taker buy quote asset volume and Ignore. First, a price volatility for each interval (or row) is calculated as a difference between the Low and High price divided by the Open price and stored in a Volatility column. Next, the dataset is sorted by ascending date, so that the last data point represents 15th June 2018 at 00:00:00. Open time, or date, column is not needed anymore as the data is sorted. As the TA indicators that will be fed into a LSTM model are calculated only from the closing price, the rest of the price columns in the dataset can be discarded. The dataset now looks as in the picture below.



After the TA indicators are calculated the values need to be normalised, so that the inputs to the LSTM are within the range -1 to 1. This is important since deep learning models do not like inputs that vary wildly.

**TA Indicators**

As mentioned in the Research review, there are 8 indicators calculated from the Close price column: Moving Average (MA), Exponential Moving Average (EMA), Double Exponential Moving Average (DEMA), Average Directional Index (ADX), Relative Strength Index (RSI), Moving Average Convergence/Divergence (MACD), Stochastic Oscillator (SO) and Williams %R. The first four are trailing indicators that show the average of price movement over some past duration – common values are 9, 21, 50, 100, 200 price points, although it depends on the size of the interval used. The next four are oscillators, which tell whether an asset is overbought or oversold and typically range between 0 and 100.

To calculate these indicators, I used Ta-Lib python wrapper for a popular C++ Technical Analysis library. All 8 indicators are merged into a single data frame with 11 columns, which represents the input data for the LSTM neural network.

I aim to finish with TA Indicators by Sunday. Next week I will focus on designing the machine learning model and playing around with different inputs.

**Data Analytics Architecture Diagram**

A screenshot of a cell phone

Description generated with very high confidence

1. https://coinmarketcap.com/ [↑](#footnote-ref-1)