

Deep Learning and Social Media in Aim to Cryptocurrency Trading.

Academic Project Work

July 15, 2019

SErgio ivan villalobos lizarazo

Master in Business Analytics and Big Data – MIP Politecnico di Milano

Supervisors: Phd. Laura Grassi, Phd. Carlotta Orsenigo.



Abstract

This Academic project studies the possibility of forecasting price movements in a specific cryptocurrency market, by implementing deep learning models, exploiting other types of features more than just the market indicators commonly used; with this purpose in mind, it is important to explore which set of features derives in a model with higher predictive power, consequently, better trading opportunities can be assessed.

The research gathers and organizes data from July 2017 until August 2019, among those features there are included Social media data (Tweets volume, Google trends scores), characteristic Technical Analysis indicators and Fundamental metrics proper of the blockchain. Two different LSTM configurations apt for multistep time series forecast are explored: sequence to vector (Vector output models) and sequence to sequence (Encoder-Decoder model) recurrent neural networks.

Models including a mix of social media data and fundamental metrics improve forecasts based on LSTM neural networks, they result in more robust and more accurate models than the ones using the typical market indicators. The use of this models can help investors to take better-informed decisions when performing day-trading in cryptocurrency markets.

Special Thanks are given to Phd. Jason Brownlee and his tutorials and books of Machine learning in <https://machinelearningmastery.com/> which help a lot not only to comprehend but also to develop and validate the models.

Table of Contents

[Chapter 1: Introduction 3](#_Toc19534764)

[1.1 Cryptocurrency 3](#_Toc19534765)

[1.1 Bitcoin (BTC) 3](#_Toc19534766)

[1.2 Bitcoin as trading instrument 4](#_Toc19534767)

[1.3 Machine learning in Trading 4](#_Toc19534768)

[1.4 Fundamental analysis vs. Technical analysis 5](#_Toc19534769)

[1.4.1 Fundamental metrics about Bitcoin 5](#_Toc19534770)

[1.4.2 Technical indicators 6](#_Toc19534771)

[1.5 Previous Work 7](#_Toc19534772)

[1.5.1 Tweets Volume: 8](#_Toc19534773)

[1.5.2 Google Trends: 8](#_Toc19534774)

[1.6 LSTM Deep Neural Networks 8](#_Toc19534775)

[1.7 Present Work 8](#_Toc19534776)

[Chapter 2: Methodology 10](#_Toc19534777)

[2.1 Data Gathering: 10](#_Toc19534778)

[GetOldTweets3 10](#_Toc19534779)

[Google Trends Data 10](#_Toc19534780)

[2.2 Data Description and Pre-Processing 10](#_Toc19534781)

[Missing Data 10](#_Toc19534782)

[Technical Analysis Library (TA-Lib) 10](#_Toc19534783)

[Ranking of features 11](#_Toc19534784)

[Final Dataset 11](#_Toc19534785)

[2.3 Multi-variate multi-step forecasting supervised models [11] 12](#_Toc19534786)

[2.4. Evaluation 13](#_Toc19534787)

[Pre-processing 13](#_Toc19534788)

[Walk forward Validation 13](#_Toc19534789)

[Training 14](#_Toc19534790)

[Tuning 14](#_Toc19534791)

[Evaluating predictions 14](#_Toc19534792)

[Residual error diagnostics 15](#_Toc19534793)

[Chapter 3: Results & Discussion 16](#_Toc19534794)

[Output Vector Tuning: 16](#_Toc19534795)

[Seq2Vec Forecasts: 17](#_Toc19534796)

[Encoder-Decoder tuning: 21](#_Toc19534797)

[Seq2Seq Forecasts: 21](#_Toc19534798)

[Main Observations & Discussion: 24](#_Toc19534799)

[Chapter 5: Conclusions & Recommendations 25](#_Toc19534800)

[Future Work 25](#_Toc19534801)

[References 26](#_Toc19534802)

# Chapter 1: Introduction

In the opening chapter of this project, concepts of cryptocurrency markets and machine learning will be exposed to acquire familiarity on the timeseries forecasting problem to be tackled, following, previous related researches raised in this same context are presented, in order to better understand the main objective of this project and why deep learning as a state-of-the-art approach for this specific subject. The second chapter describes the methodology used to answer the business question stated along with the scope of the research and for the last chapters, the reader will find detailed and discussed results of the experiments. Finally, to bring to a close, conclusions and recommendations for future work are presented.

## 1.1 Cryptocurrency

“A cryptocurrency is digital asset designed to work as a medium of exchange that uses strong cryptography to secure financial transactions, control the creation of additional units, and verify the transfer of assets. Cryptocurrencies use decentralized control as opposed to centralized electronic money and central banking systems. The decentralized control of each cryptocurrency works through a distributed ledger technology, typically a **blockchain**, that serves as a public financial transaction database”*[[1]](#footnote-1)*.

## 1.1 Bitcoin (BTC)

It wasn’t until 2009 with the creation of Bitcoin, that cryptocurrencies gained traction in the trading world. Since then, it has transformed to become the most famous cryptocurrency and the representative term for cryptocurrencies/digital currencies[[2]](#footnote-2). It was the first decentralized cryptocurrency and since then, more than 4000 other cryptocurrencies have been created (among more than 200 cryptocurrency exchanges are operating 24/7 nowadays). Although it has already been a decade since its inception, no other cryptocurrency has come close to toppling it from its dominant position4.

## 1.2 Bitcoin as a trading instrument[[3]](#footnote-3)

There are a few reasons to consider Bitcoin as a trading instrument, and sometimes even a superior one, in comparison to other asset classes (commodities, stocks and even Forex):

1. Bitcoin trading is global: It is spread across electronic exchanges worldwide and, as we noted earlier, the market is active 24/7. In other words, traders worldwide are able to take advantage of Bitcoin movement whenever it is convenient for them (outside of business hours, during weekends or holidays).
2. High volatility: Very often the cryptocurrency demonstrates movements of 3% or even larger within a single trading day. At times of exceptional volatility, these movements may exceed even 10%. Seasoned traders know that such volatility equals tremendous gains in a shorter term, if an appropriate strategy is used.
3. A more affordable instrument to trade: Fees associated with Bitcoin exchanges are far lower in comparison with traditional exchanges. In addition, Bitcoin deposits and withdrawals are usually processed in a matter of a few hours, regardless of your location on the globe. And what also appears as an advantage is the fact that personal data requirements at Bitcoin exchanges are less strict, especially if you do your deposits and withdrawals exclusively in cryptocurrency.
4. Crypto traders can operate with Bitcoin by using leverage. This combined with high volatility and constant availability provides good opportunities for private traders who usually have limited spare time and, of course, limited funds.

Why choose bitcoin over other cryptocurrencies? Bitcoin’s popularity has resulted in what is known as the “network effect” *[[4]](#footnote-4)* . It describes the increase in the value of a product or service as more users take to it. A commonly named example is Facebook, which became the world’s most widely used social media website because of network effect. With Bitcoin being the world’s most widely used cryptocurrency, the resulting network effect may “create a moat protecting it from the onslaught of other cryptocurrencies” *[[5]](#footnote-5)*.

## 1.3 Machine learning in Trading

Trading, more than merely buying and selling shares on financial markets, is about identifying localized patterns – which are often limited in time and space – and then guessing how to exploit them, the process of finding patterns is arduous and time consuming. Machine learning algorithms are basically pattern-finding machines[[6]](#footnote-6). Machine Learning offers a number of important advantages over traditional algorithmic programs. It also increases the number of markets an individual can monitor and respond to. Most importantly, they offer the ability to move from finding associations based on historical data to identifying and adapting to trends as they develop, translated in having a competitive adventage.

## 1.4 Fundamental analysis vs. Technical analysis

Fundamental analysis and technical analysis, the major schools of thought when it comes to approaching the markets, are at opposite ends of the spectrum. Both methods are used for researching and forecasting future trends in stocks. The main objective of fundamental analysis is to determine the intrinsic value of the individual asset; regardless of the asset class (bonds, commodities, securities…). If the asset’s price is trading below its intrinsic value, the investor would be inclined to buy the asset based on the fact that it’s undervalued and vice versa.

Technical analysis uses a completely different approach. It’s a trading approach designed to evaluate investment flows and trading opportunities by analyzing statistical trends. These statistical trends are gathered from various trading activities, most notably **price movement** and **volume**. Technical analysis makes no effort to determine intrinsic value, the core assumption is that all known fundamentals are factored into price.[[7]](#footnote-7) Instead, it focuses on patterns derived from price movements and charting tools, these tools are used to appraise the strength or weakness of the underlying security or asset class. Given the dramatic price fluctuations within the crypto universe, it’s virtually impossible to accurately determine the intrinsic value of any cryptocurrency, including Bitcoin. Therefore, it’s impractical to apply pure fundamental analysis if the intrinsic value is unavailable.[[8]](#footnote-8)

### 1.4.1 Fundamental metrics about Bitcoin

The first data to be gathered will be our target feature, BTC historical prices, more specifically, USD/BTC closing prices of the world’s oldest Bitcoin Exchange, and one of the largest: **Bitstamp**. Historical data is available for download through portals such as *Cryptodatadownloads.com* and *bitcoincharts.com*. Additionally, from the dedicated Crypto-data portal *Coinmetrics.com,* we can download daily public data of the following fundamental metrics related to cryptocurrencies:

#### Active Address count

The sum count of transfers that day. Transfers represent movements of native units from one ledger entity to another distinct ledger entity. Only transfers that are the result of a transaction and that have a positive (non-zero) value are counted.

#### Transfers Count

The sum count of transfers that day. Transfers represent movements of native units from one ledger entity to another distinct ledger entity. Only transfers that are the result of a transaction and that have a positive (non-zero) value are counted.

#### Average transactions counts

The mean count of native units transferred per transaction (i.e., the mean "size" of a transaction) that day. Both in native units (BTC) and dollars (USD).

#### NVT or Network Value to Transactions Ratio and Signal (Adj to 90 days)

The ratio of the network value (or market capitalization, current supply) to the 90-day moving average of the adjusted transfer value. This indicator, created by Willy Woo[[9]](#footnote-9), was an attempt to create the equivalent to a price-earnings ratio (PE Ratio) for the Bitcoin market.

### 1.4.2 Technical indicators

In general, traders use technical indicators to figure out both long & short-term price direction of an asset, it is also possible to measure anything from momentum, volume, quality of price movement, and much more. These are mathematical calculations or ‘signals’ used in technical analysis to determine what may happen next with the price of a security, commodity, stock, currency, or cryptocurrency.

Technical Indicators can be used to help you obtain a better understanding of how prices might change in the future, although there are different types: leading, lagging, oscillator and momentum indicators. The difference between a leading and a lagging indicator is that leading indicators can be used to predict what will happen to a price in the future while a lagging indicator is used to confirm or deny any trends that are or have already happened. Oscillator indicators vary back and forth between a central value, while also can warn of the end of a trend. Momentum indicators are indicators that aren't bound to a range and can be used to confirm a trend once it has started.[[10]](#footnote-10)

#### Volume Indicators:

Until now it should be clear the importance of transactions volume in technical analysis, it provides a “snapshot” picture of how many traders are actually establishing positions at various price levels[[11]](#footnote-11). Providing several clues to the underlying strength or weakness of the market.

#### RSI - Relative Strength Index[[12]](#footnote-12)

A momentum indicator that measures the magnitude of recent price changes to evaluate overbought or oversold conditions in the asset´s price, signals are considered overbought when the indicator is above 70% and oversold when the indicator is below 30%.

#### TRIX - Triple Exponential Average[[13]](#footnote-13)

An [oscillator](https://www.investopedia.com/terms/o/oscillator.asp) used to identify [oversold](https://www.investopedia.com/terms/o/oversold.asp) and overbought markets, and it can also be used as a [momentum](https://www.investopedia.com/terms/m/momentum.asp) indicator. TRIX oscillates around a zero line, when positive indicates an oversold market and increasing momentum, when below zero the opposite.

#### MACD - Moving Average Convergence/Divergence[[14]](#footnote-14)

A [trend-following](https://www.investopedia.com/terms/t/trendtrading.asp) [momentum](https://www.investopedia.com/terms/m/momentum.asp) indicator that shows the relationship between two [moving averages](https://www.investopedia.com/terms/m/movingaverage.asp) of a security’s price, it helps investors understand whether the bullish or bearish movement in the price is strengthening or weakening.

## 1.5 Previous Work

Twitter is one main social media platform for conversation and disclosure of financial information on cryptocurrencies, the initial focus of the project was to study the effect of integrating Sentiment Analysis based on Bitcoin related tweets to a predictive model, but several research projects have already tested this hypothesis in recent years, for example in the book “*Forecasting Cryptocurrency Value by Sentiment Analysis: An HPC-Oriented Survey of the State-of-the-Art in the Cloud Era*”[7] we find an extensive list of other researches giving it a try, overall, remarks the complexity and limits when considering SA in forecasting cryptocurrency; In addition, supported by the scientific report *“When Can Social Media Lead Financial Markets?”* [8] if the financial instruments attract sufficient messages, in general, the sentiments do not lead financial markets in a statistically significant way.

The 2018 research, “*Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis*”[1], concludes that the average sentiment of the messages is not a good predictor of the price of Bitcoin while the volume of the messages is. And finally, the most recent academic project work (06/2019): *“Can we commerce bitcoin using Sentiment analysis over Twitter?”*[3] has downloaded and analyzed 17 million bitcoin related tweets between August 2017 and January 2019, by applying VADER (Valence Aware Dictionary and Sentiment Reasoner). VADER is a lexicon and rule-based sentiment analysis tool that isspecifically attuned to sentiments expressed in social media, the research demonstrates that the average sentiment expressed about the crypto remains positive even when the bitcoin value falls and reveals the high correlation between Bitcoin price and tweets volume.

The present work includes two effective Social media indicators, based on those findings:

### Tweet Volume:

Count of tweets day by day, hour by hour relates to how much people are talking about bitcoin on the Twitter platform.

### Google Trends[[15]](#footnote-15):

The search trends feature provided by Google shows how frequently a given search term is entered into the search engine relative to the site’s total search volume over a given period of time. It can be used for comparative keyword research and to discover event-triggered spikes in keyword search volume.

## 1.6 LSTM Deep Neural Networks

When forecasting cryptocurrency values, a key matter is to consider if they should be classified as currencies, assets, or investment vehicles. Indeed, they can behave similarly but it is easier to consider them a hybrid asset. Traditional time series methods such as ARIMA and GARCH models are effective only when the series is stationary, which is a restricting assumption. Long Short-Term Memory networks – usually just called “LSTM” – are a special kind of Recurrent Neural Network capable of learning long-term dependencies. They provide significant progresses in being able to model the non-linearly relationships and process data with multiple dimensions in a non-linear fashion. Those methods outperform the traditional ones when applied to crypto markets as exposed in the research projects of *Sean McNally [9]*, *and Laura Alessandretti [10]*, they proved that LSTM worked best when predictions were based on longer periods of data (~50 days), not only capturing long-term dependencies but also very stable against price volatility, fit to our objective. Last but not least, the 2018 Publication of Jack Press [12] shows the ability of a LSTM Neural Network to outperform ARIMA modeling for nonlinear and non-stationary real-time data streams for cryptocurrency price prediction.

## 1.7 Present Work

The popularity of cryptocurrencies skyrocketed in 2017 due to several consecutive months of exponential growth of their market capitalization, which peaked at more than $800 billion in Jan. 2018 [10]. Accordingly, the study interval of this research covers the period since July 2017, right before this boom of the crypto-market, going through the subsequent bear-market period symbolic of most of 2018 until the last cycle of August 2019 where a new phase of a more established growing market is reported. In order to capture all of this representative movements into our predictive models and given that traditional models do not work well for very long timeseries, different configurations of LTSM neural networks will be trained, tested and tuned for both social media indicators, and the other mentioned features of different nature. In order to answer our principal question:

***“Can day-trading in cryptocurrency (bitcoin) markets be better assessed through a deep learning approach that includes, among others, social media data?”***

*And if so:*

* *Which type of data is better for building those models, which is not useful?*
* *Among the configurations of LTSM models, is one better fitted for this purpose?*

The objective would be to provide a day trader with a forecast of the next half day. This implies that the frequency of the time series will be captured hourly, predicting price movements up to 12 hours in advance.

Resolution of the timeseries analysis can be also augmented (minute by minute) or reduced (Weekly), but access to this historical data granularity is generally not open neither free. Additionally, for the scope of this project a day-trading domain is better fitted than a weekly one, given the high volatility characteristic of cryptocurrencies.

# Chapter 2: Methodology

## 2.1 Data Gathering:

The present work uses the aforementioned database published on Kaggle.com[3] extracting the hourly count of tweets since 08/2017 until 01/2019 , For the missing periods of 07/2017 and from 02/2019 until 08/2019, data has been acquired utilizing the same method that the Kaggle DB used through the GetOldTweets3 python API**.**

GetOldTweets3 [[16]](#footnote-16)

Twitter’s Official API has the limitation of time constraints, it’s not possible to get older tweets than a week, so this library mimics the user’s behavior of Twitter search on browsers and scrolls down until it collects the deepest, oldest tweet in your search.

Google Trends Data

Google provides the relative search volume for a keyword indexed between 0 (minimum relative search interest) and 100 (maximum search interest) within the selected time range. It is important to point out that the data is always indexed for the selected period of time. Thanks to the effort of *Samantha Molnar[4],* I could adapt part of her code to correctly obtain the hourly Google trends data over the two years period desired.

## 2.2 Data Description and Pre-Processing

### Missing Data

The 4.770.070 collected tweets collected for the commented periods have been parsed as hourly time series with the help of the Pandas Python library. There is a total of 786 hours that the API has not been able to collect due to deleted or no published bitcoin tweets. The hourly data provided from *Cryptodatadownload.com* has some incomplete price and volume values during the dates between 07/2017 and 11/2017. In this case we have completed the missing data from *Bitcoincharts.com* targeting also the Bitstamp BTC/USD trading pair to keep consistency. Not collected or zero google scores have been identified, corrected, and imputed using the library “impute TS”[[17]](#footnote-17) in R-studio by applying Spline interpolation.

### Technical Analysis Library (TA-Lib)

TA-Lib is widely used by trading software developers requiring to perform technical analysis of financial market data. Including 150+ indicators such as ADX, MACD, RSI, Stochastic, Bollinger Bands, etc., it is an open-source API for C/C++, Java, Perl, Python and will be used to calculate the technical indicators mentioned in *1.4.2 Technical indicators.*

### Ranking of features

Each category of features will be tested first separately, and then it is chosen one best set of features, trying a more accurate model since not all features may be useful. Feature selection is applied calculating the Maximal Information Coefficient (MIC) of the R-library “Minerva”. And the correlation score (Python Pandas) of each feature against the target variable. NVT signal and avg. transaction counts in USD are chosen over the avg. transaction counts in BTC units, and the NVT Signal is preferred over the NVT Ratio, as they represent the same info with better MIC score.

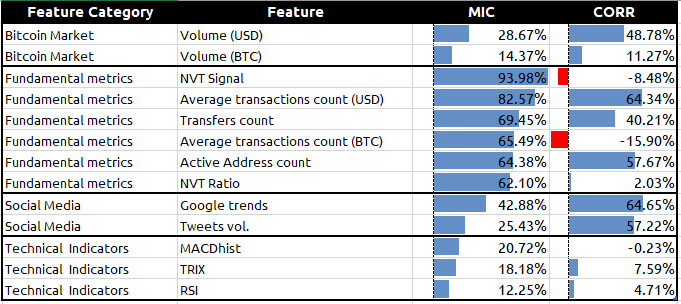


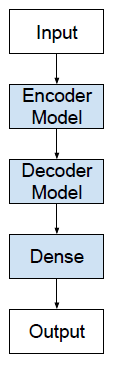
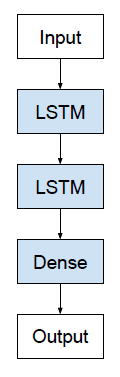
Table 1. MIC and Correlation scores of our features against the target variable.

### Final Dataset

The total set of 18.757 hours is compound of 11 features/dimensions. It is important to note that the periodicity for the fundamental metrics is not available on an hourly basis, but daily, and they will be placed for the 24 corresponding hours on the dataset. Some empty values appear at the beginning of technical indicators as a result of their calculation.

## 2.3 Multi-variate multi-step forecasting supervised models [11]

Sequence prediction is different to other types of supervised learning problems. The sequence imposes an order on the observations that must be preserved when training models and making predictions. If the input and output sequences are a time series, then the problem may be referred to as multi-step time series forecasting**.** This is where you have one or multiple series with multiple input time steps and wish to predict multiple time steps beyond one or more of the input sequences. This can be implemented as a many-to-many model. Furthermore, when you have multiple series with multiple input time steps and wish to predict one-time step beyond one or more of the input sequences, it is called as a Multivariate Time Series Forecasting and it can be implemented as a many-to-one model, each series is just another input feature:



Stacked LTSM (Seq2Vec)

Encoder Decoder (Seq2Seq) Architechture

**Many-to-one:**

**Many-to-Many:**

Figure 1. Recurrent Neural networks Architectures (outside) to models (inside)

For the many-to-one model, we will use a common architecture of RNN called Stacked LSTM. This may be confusing because we clearly intend to predict a sequence of output time steps and not just the next closing price of BTC. The reason that this is a many-to-one prediction problem is because the model will not predict the output time steps piecewise; the whole prediction will be produced at once to a single output vector, also known as a Sequence to vector (seq2vec). There is a more challenging type of sequence prediction problem as it takes a sequence as input and requires a sequence prediction as output: sequence-to-sequence prediction problems (seq2seq). One design concern that makes these problems challenging is that the length of the input and output sequences may vary. A common architecture that is apt for this approach is the Encoder-Decoder model as the model will not output a vector sequence directly. Instead, the model will be comprised of two sub models: the encoder to read and encode the input sequence and the decoder that will read the encoded input sequence and make a one-step prediction for each element in the output sequence. The difference is subtle, as in practice both approaches do in fact predict a sequence output. The important difference is that an LSTM model is used in the decoder, allowing it to both know what was predicted for the prior hour in the sequence and accumulate internal state while outputting the sequence.

Keras and SciKit libraries in python allow to implement easily this type of RNN upon TensorFlow™.

## 2.4. Evaluation

### Pre-processing

Some transformations are required so Keras RNN can process the data to build and fit the models:

* “NaN” observations are eliminated, corresponding to 4.19% of the total set.
* 80% of samples are used for training and 20% (almost 5 months for testing).
* Standardization with the RobustScaler from sci-kit-learn library is conducted on training set and then applied to the test & training set.
* Sequences must be reframed as supervised learning problems. In other words, they need to be divided into input () and output pairs () and then as making a prediction based on a function of the current and previous timesteps:

### Walk forward Validation

By doing this sequence reframing and deciding the amount of previous inputs to use to get the next output, we are setting a fixed size window of “n” previous time steps and moving it one step at a time assigned to the corresponding prediction (t+1). Because this methodology involves moving along the time series one-time step at a time, it is often called Walk Forward Validation or Rolling Forecast.

To better understand which category of features can actually improve the predictive capacity of models, different subsets will be tested corresponding to the original market indicators in addition to the features of each category:

|  |  |  |
| --- | --- | --- |
| **Name of Data subset** | **Dimensions** | **Features included** |
| **Basic\_DS** | 3 | Only BTC market indicators (3) |
| **Social\_DS** | 5 | BTC market ind. (3) + Social Media (2) |
| **Fundamental \_DS** | 6 | BTC market ind. (3) + Fundamental metrics (3) |
| **Technical\_DS** | 6 | BTC market ind. (3) + Technical ind. (3) |
| **Ranked\_DS** | 7 | BTC market ind. (3) +Best ranked features:  NVT signal + G.trends + Avg Transaction USD+Tweets |

Table 2. Subsets by Category of features

### Training

* Training will be done over multiple epochs (complete passes through the dataset) including a type of regularization to restrict overfitting called “early stopping” included in Keras as a callback function that monitors & stops the fitting process when error starts to increase.
* Loss function = mean squared error (MSE).
* Optimizer: Adaptive moment estimation Algorithm (Adam)
* Activation function = tanh

### Tuning

Some other hyperparameters are tested for each case of [subset,model] before building their definitive models:

* Encoder-decoder: batch size and number of neurons.
* Vector output: batch size, number of neurons, dropout and number of layers.

### Evaluating predictions

In order to determine how near the fitted values are to the actual ones, the predictions must first be obtained and rescaled to be correctly compared with the actual values. Next, the chosen evaluation metrics are the Root Mean Squared Error (RMSE) and the Mean Absolute Percent Error (MAPE), as MAPE is scale-independent and represents the ratio of error to actual values as a percent.

### Residual error diagnostics

After a forecasting model has been fit, it is important to assess how well it captures patterns. In other words, evaluating whether the model properly fits the timeseries through their residuals:

Residual errors are expected to behave as white noise, as they represent what cannot be captured by the model, by means of the following properties:

* The residuals are uncorrelated. (Acf=0) If there are correlations between residuals, then there is information left in the residuals which should be used in computing forecasts.
* The residuals have a mean equal to zero. If the residuals have a mean other than zero, then the forecasts are biased.

# Chapter 3: Results & Discussion

### Output Vector Tuning:

Test Mean Squared Error results of different hyperparameters for a 2 LTSM Stacked layers:

Figure 2. Seq2Vec Test MSE results of each subset for different Batch sizes

Figure 3.Seq2Vec Test MSE results of each subset for different number of neurons in each hidden layer

Figure 4. Test MSE results of each subset applying different Drop Out in each hidden layer

After evaluating different hyperparameters, for this architecture, the higher error in most of cases is reached with the Basic\_DS. On the other hand, it seems that the fitting process scores a lower MSE error with the Fundamental Metrics Dataset.

In general, a lower error is accomplished for batch size of 64 samples, 50 neurons in each layer with a drop out of 0.2. Thus, these will be the set of hyperparameters chosen for the final Seq2Vec Models.

### Seq2Vec Forecasts:

**Output-Vector of 2 Stacked LTSM Layers,** 50 neurons per layer, 0.2 drop out and batch size 64 performances:

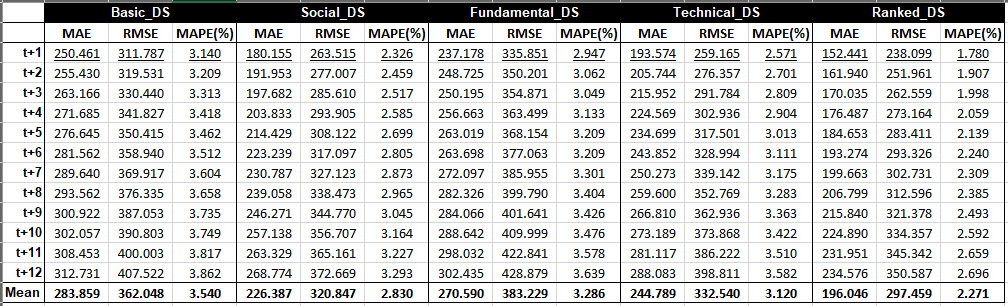


Table 3. Seq2Vec Error Scores for each timestep by Dataset – 2 Stacked LTSM Layers

Within this architecture the relation observed among the timestep predicted and the error is directly proportional. This means that the further the prediction goes the more inaccurate it is.

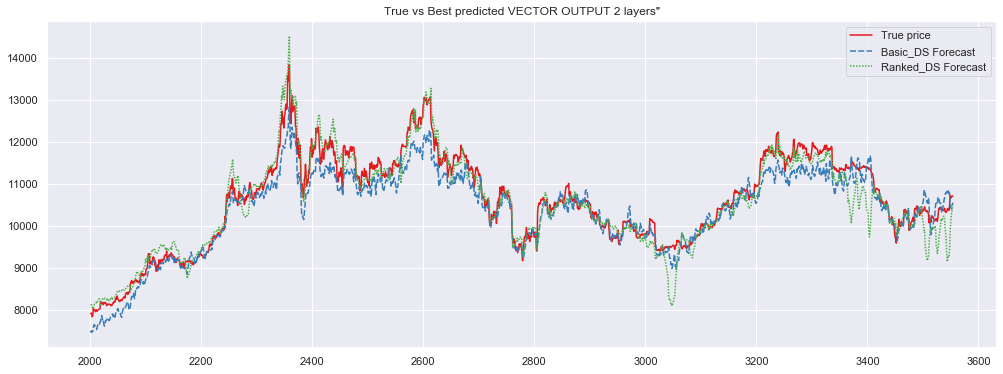


Figure 5. Seq2Vec prices forecasts – 2 Stacked LTSM Layers

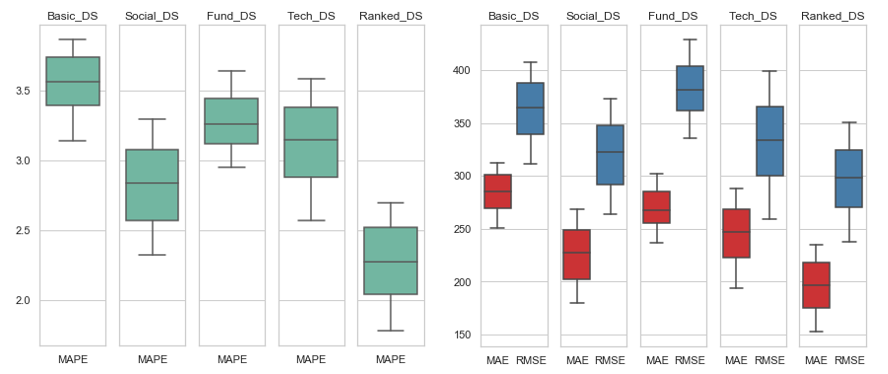
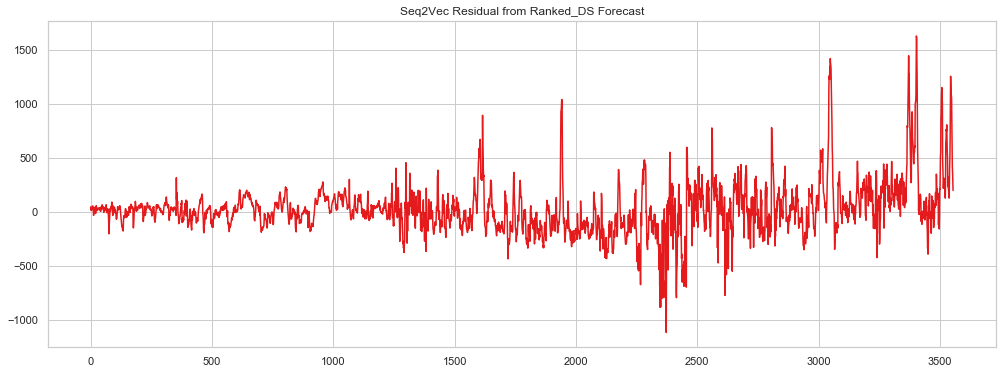


Figure 6. Seq2Vec Mean absolute percentage error Boxplots (left) - 2 Stacked LTSM Layers  
 Seq2Vec Mean absolute, Root mean squared error Boxplots (right) - 2 Stacked LTSM Layers

Although there was indication previously in the fitting process that the fundamental features would possibly lead to a better forecast, the RMSE of this experiment shows a higher error than the basic feature model for this LTSM architecture.

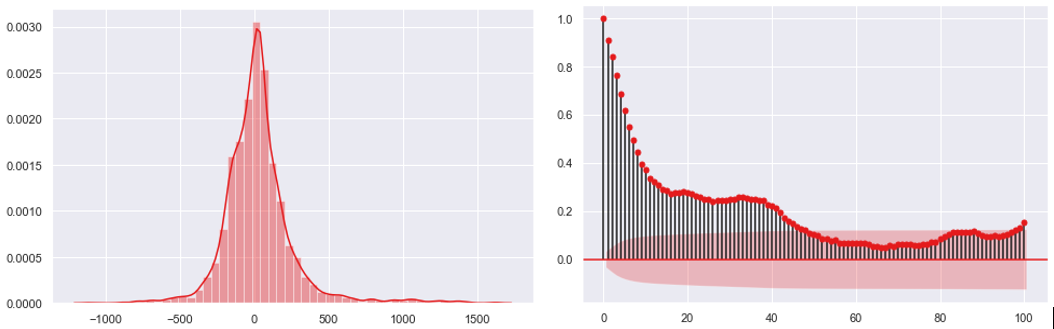


Figure 7. Seq2Vec Residual errors of Ranked DS forecast, Density plot(left) and Autocorrelation plot (right) -2 Stacked LTSM Layers

The variations of the predictions are more accurately forecasted at the beginning than at the last part of the timeseries, observable also in the high peaks of the residuals and the ACF plot.

**Output-Vector of 3 Stacked LTSM Layers,** 50 neurons per layer, 0.2 drop out and batch size 64 performances:

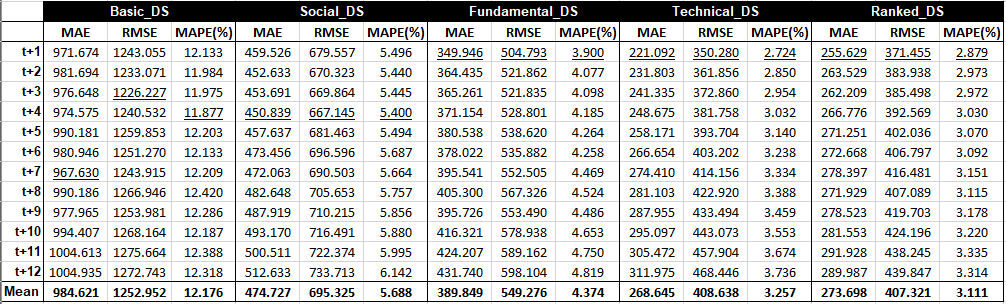


Table 4. Seq2Vec Error Scores for each timestep by Dataset – 3 Stacked LTSM Layers

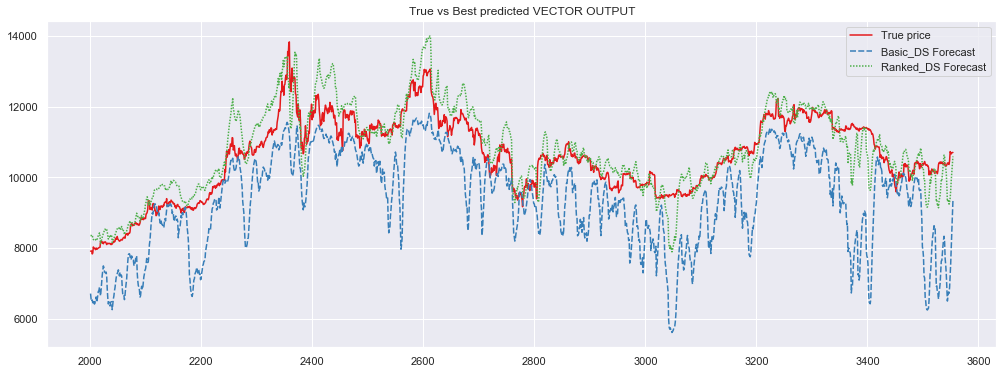


Figure 8. Seq2Vec prices forecasts – 3 Stacked LTSM Layers

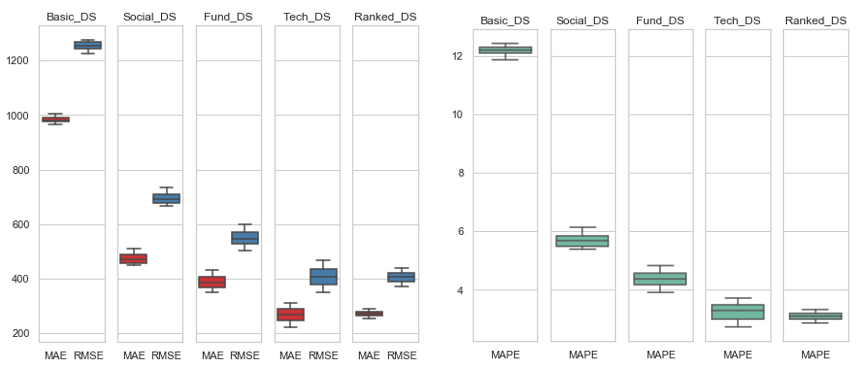
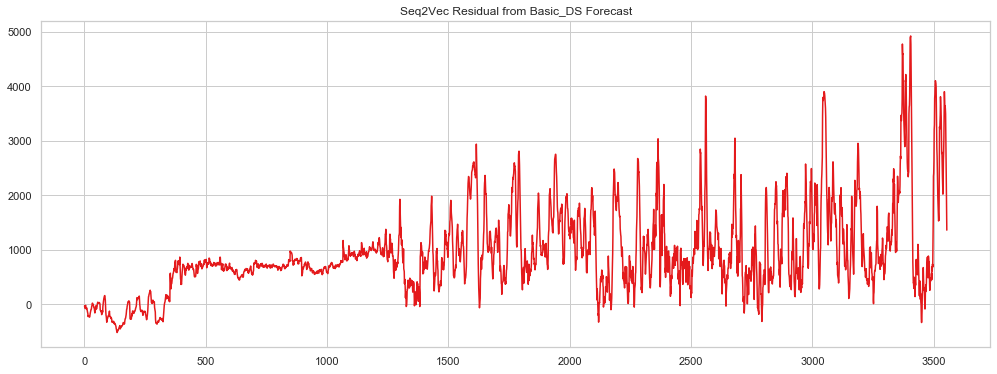


Figure 9. Seq2Vec Mean absolute percentage error Boxplots (right) - 3 Stacked LTSM Layers  
 Seq2Vec Mean absolute, Root mean squared error Boxplots (left) - 3 Stacked LTSM Layers



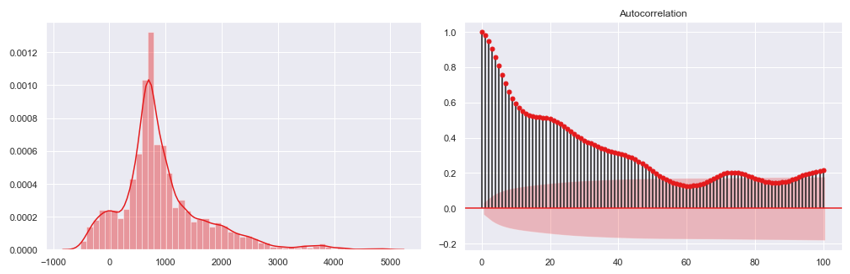


Figure 10. Seq2Vec Residual errors of Basic\_DS forecast, Density plot(left) and Autocorrelation plot (right) -3 Stacked LTSM Layers

It is evident the forecasts are less accurate by adding another LTSM layer—particularly for the basic DS. The density plot and the time series plot demonstrate that the forecast is biased.

### Encoder-Decoder tuning:

Test Mean Squared Error results of different Batch sizes and Number of neurons for the Encoder-Decoder configuration:

Figure 11. Seq2Seq Test MSE results of each subset for different Batch sizes

Figure 12. Seq2Seq Test MSE results of each subset for different number of neurons in hidden layers

### Seq2Seq Forecasts:

**Encoder-Decoder** configuration with the best performed batch size and number of neurons for each Dataset, for this specific experiment:

Basic\_DS: [ batch size = 128, Neuron number = 250 ]

Social\_DS: [ batch size = 64, Neuron number = 50 ]

Funamental\_DS: [ batch size = 128, Neuron number = 250 ]

Technical\_DS: [ batch size = 128, Neuron number = 50 ]

Ranked\_DS: [ batch size = 64, Neuron number = 250 ]

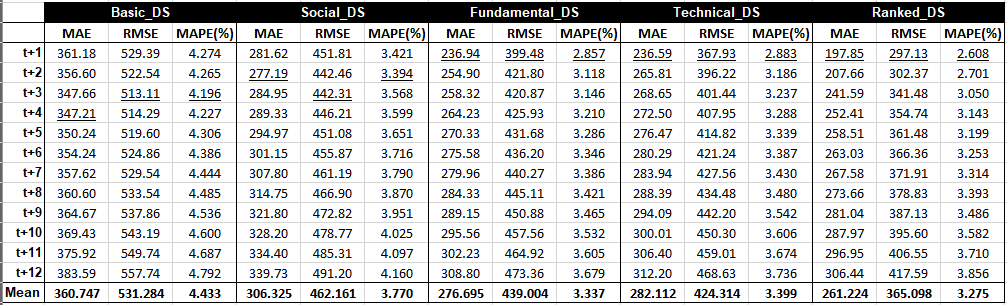


Table 5.Seq2Seq Error Scores for each timestep by Dataset

With the Seq2Seq forecast not always the most accurate prediction is for the next immediate timestep (hour).

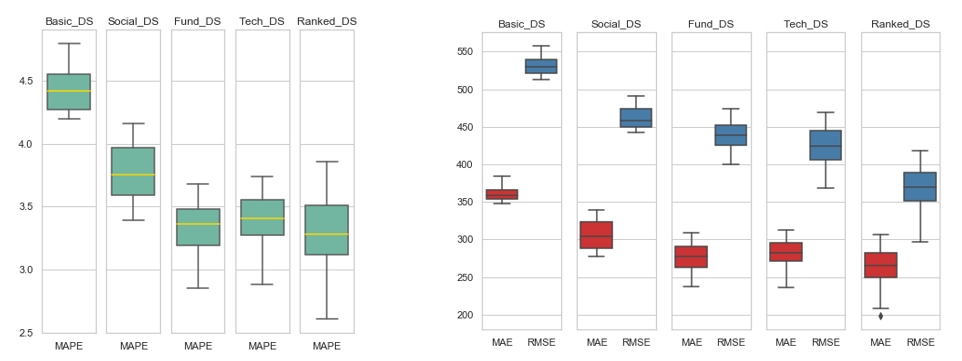


Figure 13. Seq2Seq Mean absolute percentage error Boxplots (left)  
 Seq2Seq Mean absolute, root mean squared error Boxplots (right)

The best predictive models are the ones with the blockchain fundamental metrics and the best ranked features:

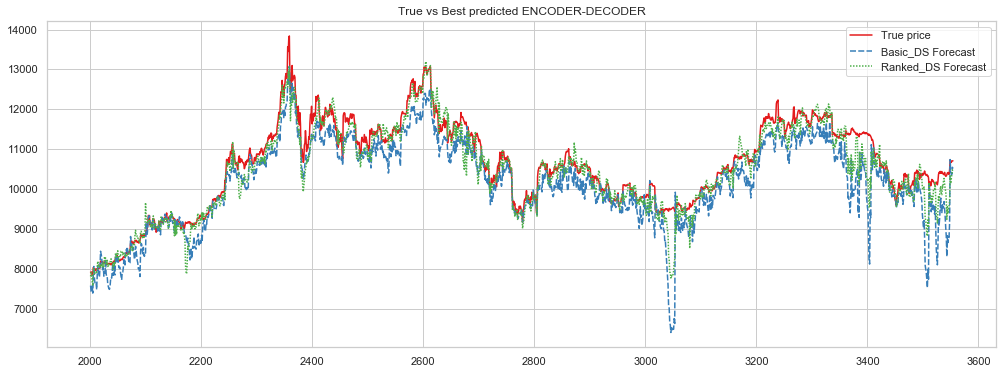
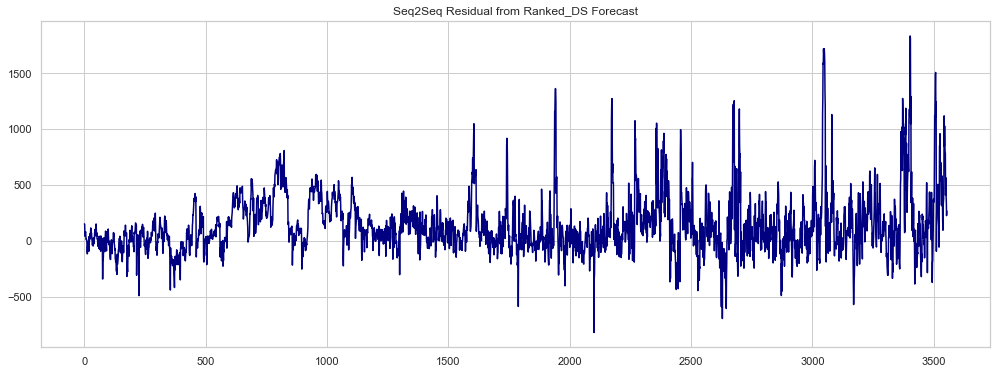


Figure 14.Seq2Seq prices forecast for the Basic dataset (greatest error) and the Fundamental metrics (lowest Error)



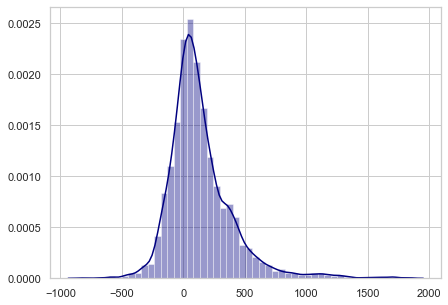
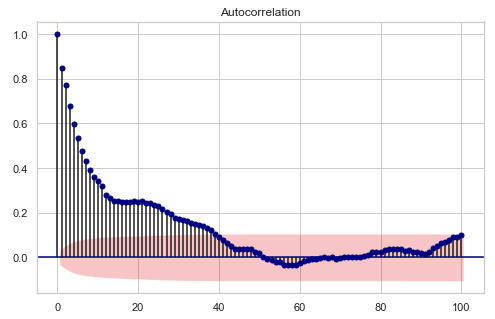
 

Figure 15.Seq2Seq Residual errors form best prediction (upside) their Density plot(left) and Autocorrelation plot (right)

The Ranked\_DS Model residuals have almost zero mean and constant variance, seeming to be normally distributed. However, the ACF plot shows that the model still misses opportunities to capture patterns of the timeseries despite the improvement obtained when compared to the other sets of features.

### Main Observations & Discussion:

* Both Seq2Vec and Seq2Seq Forecasts present similar results in terms of accuracy (RMSE and MAPE).
* The models with better predictive power among the experiments are the sets composed of a mix of best ranked features according the MIC metric. The most accurate model with a prediction error of 2.27% is accomplished with the 2 Stacked LTSM layers configuration.
* Volatility of the predictions is always attenuated by integrating other type of features (any other category).
* The sets containing just one category of features (tech, fund and social) scored very similar in between them, improving the predictive power of models just with the typical market indicators. In each configuration, each of them scored better than the other two meaning this assumption should not be generalized.
* Despite technical indicators not having very good correlation or MIC scores against our target variable, they are still valid features to be used for predicting.
* The basic set of features resulted in high sensitivity with respect to some hyperparameters, particularly the number of neurons used in the hidden layers. This led to the model suffering from high variance. On the contrary, the rest of the models were more robust against these changes.

# Chapter 5: Conclusions & Recommendations

Regarding our original questions, the present work confirms that it is possible to assess better trading decision in a day trading context with models that use other types of data. Models including a mix of social media indicators and fundamental metrics from the blockchain improve forecasts done through LSTM, result in more robust and more accurate models than the ones using the typical market indicators. There is not sufficient evidence to determine which configuration of LSTM (Seq2Vec or Seq2Seq) is better fit for the original purpose.

The internet is flooded with very simple approaches of timeseries forecasting approaches with deep learning (particularly the one-step prediction with neural networks) for achieving a good forecast model. Caution must be taken when deciding on a validation method for analyzing forecast results; in our case, a walk forward validation (rolling window) was used.

### Future Work

Given the lengthy time used for testing due to my reduced power of computation (especially for the encoder decoder models), few experiments were done to obtain the adequate hyperparameters for each subset. To have stronger statistic assurance, more experiments and ANOVA testing should be done to define with more certainty the adequate hyperparameters for each configuration.

Technical Indicators should be used and interpreted with caution as their definition confirms they are not flawless strategies by themselves. They depend a lot, among other factors, on the periodicity of the analysis. Expert traders use a mix of indicators to derive signals of buying or selling points. Selecting different indicators than the ones used in this project and translating them into categorical features instead is a strategy worth investigating.

Day Traders also analyze 4-hours and 8-hours periods--indicators famous among day trading. This analysis can be repeated by changing the periodicity of the data.

Making the target feature (BTC closing prices) stationary by differentiating it, as suggested by PhD. Jason Brownlee [11], results in a volatility analysis. And by then reversing the differentiation transformation, it is possible to obtain more accurate forecasts.

To take greater advantage of the Encoder-decoder benefits, tests using different input time steps (for example using only 6 or 4 hours as input to check how far in terms of accuracy) can be predictions of the same output (12 hours a head).

# References

1. Hassani, Hossein; Huang, Xu; Silva, Emmanuel. (2018). "*Big-Crypto: Big Data, Blockchain and Cryptocurrency*." Big Data Cogn. Comput. 2, no. 4: 34. <https://www.mdpi.com/2504-2289/2/4/34>
2. Abraham, Jethin; Higdon, Daniel; Nelson, John; and Ibarra, Juan (2018) "*Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis,"* SMU Data Science Review: Vol. 1 : No. 3 , Article 1. <https://scholar.smu.edu/datasciencereview/vol1/iss3/1>
3. Badiola Ramos, Jaime (2019) *“¿Podemos comerciar Bitcoin usando análisis de sentimiento sobre Twitter?*” - Universidad Pontificia de Comillas TFG
4. Bitcoin 17.7 million Tweets and Price Dataset –License CC0: Public Domain <https://www.kaggle.com/jaimebadiola/bitcoin-tweets-and-price>
5. Molnar, Samanta (2017) “Hacking Google Trends” <http://samanthamolnar.me/personal/2017/05/02/hacking-google-trends.html>
6. Colianni, S., Rosales, S.M., Signorotti, M.: *“Algorithmic trading of cryptocurrencybased on twitter sentiment analysis”* (2015) <http://cs229.stanford.edu/proj2015/029_report.pdf>
7. Zamuda, Aleš; Crescimanna, Vincenzo; Burguillo, Juan C.; (2019) *“Forecasting Cryptocurrency Value by Sentiment Analysis: An HPC-Oriented Survey of the State-of-the-Art in the Cloud Era”* Chapter 3- Springer International Publishing -<https://doi.org/10.1007/978-3-030-16272-6_12>
8. Zheludev, Ily; Smith, Robert; Aste, Tomaso (2014) *“When Can Social Media Lead Financial Markets?”* - Scientific Reports vol. **4**, Article number: 4213 <https://doi.org/10.1038/srep04213>
9. McNally ,Sean; “*Predicting the Price of Bitcoin Using Machine Learning*,” (2016) - National College of Ireland MSc research project <http://trap.ncirl.ie/2496/1/seanmcnally.pdf>
10. Alessandretti, Laura; ElBahrawy, A.; Aiello, L.M.; Baronchelli, “*Anticipating Cryptocurrency Prices Using Machine Learning*”. Volume 2018, Article ID 8983590 <https://doi.org/10.1155/2018/8983590>
11. Brownlee, Jason; *“Long Short-Term Memory Networks With Python: Develop Sequence Prediction Models with Deep Learning”* (2017). First Edition.
12. Press,Jack; *“LSTM Online Training and Prediction: Non-Stationary Real Time Data Stream Forecasting”* (2018) - Wayne State University. <https://www.researchgate.net/publication/328228359>

1. https://en.wikipedia.org/wiki/Cryptocurrency [↑](#footnote-ref-1)
2. Mourdoukoutas, P. Bitcoin, Ethereum and Litecoin Are the Most Popular Cryptocurrency Investments

   Among Millennials. 2018. https://www.forbes.com/sites/panosmourdoukoutas/2018/03/25/bitcoin-ethereum-and-litecoin-are-the-most-popular-cryptocurrency-investments-amongmillennials/#3e55870576dd [↑](#footnote-ref-2)
3. http://www.binarytribune.com/bitcoin-guide/why-should-you-choose-bitcoin-as-a-trading-instrument/ [↑](#footnote-ref-3)
4. https://thecryptograph.com/analysis/the-bitcoin-effect-why-bitcoin-remains-dominant/ [↑](#footnote-ref-4)
5. https://thecryptograph.com/analysis/the-bitcoin-effect-why-bitcoin-remains-dominant/ [↑](#footnote-ref-5)
6. https://www.datarevenue.com/en/usecases/machine-learning-for-trading [↑](#footnote-ref-6)
7. https://www.investopedia.com/ask/answers/what-are-leading-lagging-and-coincident-indicators/ [↑](#footnote-ref-7)
8. https://digitexfutures.com/news/a-look-at-the-best-cryptocurrency-trading-strategies/ [↑](#footnote-ref-8)
9. https://woobull.com/introducing-nvt-ratio-bitcoins-pe-ratio-use-it-to-detect-bubbles/ [↑](#footnote-ref-9)
10. https://www.nasdaq.com/forex/education/indicators-overview.aspx [↑](#footnote-ref-10)
11. https://cryptopotato.com/crypto-trading-academy-why-volume-matters-in-trading-bitcoin-crypto/ [↑](#footnote-ref-11)
12. https://www.investopedia.com/terms/r/rsi.asp [↑](#footnote-ref-12)
13. https://www.investopedia.com/articles/technical/02/092402.asp [↑](#footnote-ref-13)
14. https://www.investopedia.com/terms/m/macd.asp [↑](#footnote-ref-14)
15. https://www.wordstream.com/google-trends [↑](#footnote-ref-15)
16. https://pypi.org/project/GetOldTweets3/ [↑](#footnote-ref-16)
17. https://cran.r-project.org/web/packages/imputeTS/imputeTS.pdf [↑](#footnote-ref-17)