NOVA IMS Information Management School

# BUSINESS CASES WITH DATA SCIENCE

## MASTER DEGREE PROGRAM IN DATA SCIENCE AND ADVANCED ANALYTICS – MAJOR IN BUSINESS ANALYTICS

**Business Case 4 – Cryptocurrency Value Prediction**


Group R

Andreia Bastos, number: 20210604

João Silva, number: 20211014

Pauline Richard, number: 20211019

Tiago Quaresma, number: 20210766


May, 2022


NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

# INDEX

## Index of Figures

# 1. INTRODUCTION

The purpose of the project is to develop a predictive model that will allow Investments4Some, a private hedge fund management firm, to forecast and predict 10 cryptocurrencies daily. With a low associated error, the model should be able to identify and predict trends for the following day. To accomplish it, datasets with daily values from the start of each currency to April 25th, 2022 were given. To achieve the expected output, we will be following the Crisp-DM method.

# 2. BUSINESS AND DATA UNDERSTANDING

## 2.1. Business Context and Objectives

Investments4Some is a long-standing Portuguese, privately held hedge funds management firm. They use traditional statistical methods and financial indicators to measure the quality of their portfolios. A few years ago, the company begun to explore the usage of Machine Learning models for market price forecasting (a technique used to predict future events based on historical data) but given the lack of maturity of the company in Machine Learning, their newly created department is failing to successfully develop these models and bring them into production.
The company is aware of the lack of sophistication of the average investor and is aware of the potential of Machine Learning methods to anticipate market trends and increase the expected returns of their investments. And for that reason, they have solicited a good forecasting model, to predict market prices and anticipate trends more accurately.

## 2.2. Data Description

As we previously indicated, we have been handed datasets composed of daily values from the start of each coin to April 25th, 2022. In Figure 1 are the coins valued in terms of USD and the datasets with respective descriptions.

# 3. DATA PREPARATION AND TRANSFORMATION

We started by creating a Data Frame for each coin that aggregates all the existing information from the various datasets as well as the Date that we convert to datetime. As a result, we ended up with 10 Data Frames. Then, while looking through the data frames, we discovered that the cryptocurrency AVAX data frame contains records before the launch date, which we decided to delete.

Furthermore, we decided to analyze the relations between the cryptocurrencies in Figure 2. As we can observe, there appears to exist correlations between all of them, with the lowest being between Axie and Chainlink (**0.11**) and one of the highest being between Ethereum and Bitcoin (**0.93**). Moreover, when it comes to Bitcoin, the top correlations lie with Ethereum and Chainlink (**0.93**), followed by Cardano (**0.89**) and Cosmos (**0.88**). On the other hand, the least correlated with Bitcoin is Axie (**0.54**), followed by Avalanche (**0.58**). We found that an individual analysis of Bitcoin was necessary since it is the first and most famous cryptocurrency.

# 4. TECHNICAL INDICATORS AND ANALYSIS

## 4.1. Technical Indicators

Before we can predict the future prices of the 10 cryptocurrencies, it is important to get some statistical trends and insights from the historical trading activity. By identifying patterns in the past price, volume, and interest movements, it will help us improve our predictions and have a better understanding of the market and its opportunities.  We are therefore using 3 different technical oscillator indicators to analyze the buy and sell signals and come up with trading ideas and potential decisions.

For the following analysis, we will focus our example on only one type of cryptocurrency: the Bitcoins (BTC), but they can all be applied similarly to the 9 others. We will also focus our interpretations of the indicator on a 1-year period (365 days), from April 25$^{th}$, 2021 to April 25$^{th}$, 2022.

### 4.1.1 Candlestick graph with Range Slider

Before calculating the technical indicators, we visualized the price variations over time, using a Candlestick graph with Range Slider. We used the opening prices, high, low and closing prices of Bitcoins, to build this graph. We can see on Figure 3, the price actions between June 2017 and the 26th of April 2022. The red candlesticks represent negative price changes of Bitcoins, while the green ones show a positive change.

Looking at the variation of the Bitcoins prices since 2017, we can see it reach a peak on Dember 15, with 19,345 USD and then quite a steady movement with small bursts activity for the next 2 years. In 2020, it is the beginning of the COVID 19 pandemic with the economy being radically slow down, and in parallel, we can observe the activity increasing back again, with the year starting at 7,339 USD and finishing at 29,224 in Dec 2020, meaning a 398% increase of the Bitcoin's price.

The next highest peak is the recorded in April 2021 with a value of 64,863 USD, matching the release of the new currency exchange, Coinbase. It then experienced a new dropdown during the summer 2021, and a new high in November 2021 at 68,917 USD, that matches the announcement of Tesla acquiring 1,5 billion US dollars' worth of the digital coin. More recently, we can see the price fell again to 36,366 USD in January 2022, and since then, it keeps fluctuating between 37,000 USD and 47,000.

### 4.1.2 Relative Strength Index (RSI)

The RSI, or Relative Strength Index, is one of the most common oscillator indicators, with the MACD and the Stochastic Oscillator. It is named this because it is a line that oscillates between two extremes: 0 and 100 and is used by traders to spot dangerous signals of the market condition.

We analyzed the RSI during the period mentioned above (Figure 4). First, we can see that the threshold of 70 has been reached multiple times, with the maximum reached on July 30 with 75, and on October 15 with 74, which matches the 2 bursts in Bitcoins' price observed in Figure 3. When the RSI reached 70 it is sending an overbought signal, meaning the gain on this asset might exceed its actual value and the price should be readjusted.

However, for example, in September 2021, we could observe a drop in prices with an RSI level of 34. Even though the level below 30 has not been reached, if it did, it would send a signal that the asset might be oversold, meaning it is traded at a price below its value and is ready for a rebound.

### 4.1.3 Moving Average Convergence Divergence (MACD)

The MACD, or Moving Average Convergence/Divergence, is an indicator that helps traders to identify the market trend direction and its momentum. The histogram represents the difference between 2 lines: the MACD line, which measures the distance between two different moving averages (26-period EMA from 12-period EMA), and the signal line, which indicates changes in price momentum (9-period EMA). We call Convergence when the two moving averages are moving towards each other, and Divergence when they are moving away from each other.

In Figure 5, we are analyzing the same period as above. First, looking at the MACD line (in blue), we can see it is below 0 only before end of July 2021, which matches the beginning of the upward phase of the price around July 21st, and after mid-November 2021, which we could also observe in the variation of the RSI. Otherwise, the 4 months period in-between was mostly above 0, indicating the market was entering a short bullish period, with a rise in prices and a boost of confidence of Investors (because of Tesla and Coinbase).

Moreover, we can also complete our analysis by looking at the behavior of the signal line (in red). When the MACD line is below the signal line, traders can interpret it as a buy signal, while when it is above it, traders should interpret it as a sell signal. Here on the graph, we can see that the sell signal obviously matches the peak of the upward trend in August and October.

### 4.1.4 Stochastic Oscillator

The Stochastic Oscillator is the perfect complement to our two other indicators. It is also used by traders to predict the overall price trend with overbought (values above 80) or oversold signals (values below 20). It is plotted between 0 and 100. The graph contains two different lines, %K which reflect the actual value of the indicator and %D, considered as the "slow stochastic indicator," because it represents a 3-day moving average of %K.

Looking at the smaller period (from July to November 2021), we can observe in Figure 6 four different buy signals. Each of them corresponds to the moment where the indicator line drops below 20 and then rises back immediately after. Since it is a period where we expect the price to rise (as seen previously with RSI and MACD), looking at when the line moves above 80 is not significant, since it is the expected movement. However, we can look at the beginning of July and September, where we could observe a downtrend. For those periods, looking at when the line goes above 80 and then drops after, is significant to identify sell signal. On our graph, we can observe two of these potential short trades. The sell and buy signals identified are manually indicated in Figure 7.

## 4.2. Technical Analysis

Technical analysts try to predict the future price movements using the past trading patterns, analyzed with indicators. Based on our analysis above, we can see that the basic monthly trends of the market are well reflected using our 3 indicators. However, we can also conclude that trading is a daily activity and changes can be observed from one day to another, and it is especially true for the Bitcoins price which is a very volatile market. Therefore, investors should always have an eye open to specific signals but also to external factors that might affect market behaviors.

## 5. IMPLEMENTATION OF MACHINE LEARNING MODELS

Aside from technical analysis, one of the company's goals was to develop a predictive model that could properly anticipate future crypto values. To achieve so, a variety of models were used in the forecasting and time-series analysis, more particularly two commonly used RNNs, LSTM (our benchmark model) and GRU (a more compact and faster version), as well as the XGBoost model. For the RNNs, we used both single and multi-variable input shapes, while the XGBoost was only given the 'close' variable.

To build the models, the first step was to obtain the last 365 days of each coin's Dataframe to use as a base for predicting. We considered one year a decent period to base our predictions on due to the substantial change in prices in earlier years. In the next step we implemented three functions, *df_to_X_y*, that creates our sliding window, *choose_sequential_model*, that runs the RNNs depending on the 'model_name' parameter, and finally *run_fit_model* that includes the two previous functions.

Our final function *coin_predictions*, given data of a coin, the model's name, and passing the one of the 4 flags (CLOSE, OHLC, IND and OHLC-IND), makes our final prediction. If the parameter metric_details is set to 1, it displays information on the Test's RMSE, MAE and MSE, to be further used as performance metrics. This function returns our forecasted dataset, based on the previous 20 days (about 3 weeks) as well as the original dataset, that will be used to add a new row with the predictions, predict the following day, and so on.

When dealing with the features to use, we immediately discarded "volume" and "adj_close", and added 'rsi' and 'macd' from the Technical Analysis and tried 4 different approaches on how to input our data to feed the RNNs:

1. Feeding our model only the "close" variable (CLOSE);
2. Using the OHLC, although highly correlated with 'close' the other 3 variables could show better results (OHLC);
3. Using "close" together with the two financial indicators "rsi" and "macd" (IND);
4. Using OHLC plus the Indicator variables (OHLC-IND).

As mentioned before, we used LSTM as our baseline model, but we also used Bitcoin as our baseline coin to test, since in the digital currency space, it is common for many coins and tokens to move in similar patterns. When bitcoin (BTC), the largest cryptocurrency by market cap, goes up, other digital tokens tend to increase in value as well. When BTC declines, it is likely that other players in the space will drop at the same time. For this reason, we chose to test our model in Bitcoin and apply the best results into other coins.

## 6. EVALUATION OF THE RESULTS

For the evaluation metric we chose the Root-mean-square error (RMSE). Since the errors are squared before being averaged, the RMSE gives more weight to higher errors. Meaning that, when these are undesired, the RMSE should be more useful. Our evaluation was based on the results of applying our predictive models in bitcoin, and for each of our four feature combination options we obtained the results on Figure 8.

When choosing our winning model, we tried to do a balance between the value that was closest to the price of Bitcoin on the 26th of April and the best RMSE value we could get. Taking this into account, we ended up choosing the GRU model with 600 epochs, and an RMSE of around 0.032, an excellent value since it is close to 0, and using the IND feature configuration explained above. In Figure 9 we can also see how other performance metrics behave, showing an expected similar pattern along the 600 when compared to RMSE.
This was an expected result, since GRU is a RNN that works better with smaller datasets (<1000 rows), which was our case (slice of 365 rows), when compared to LSTM.

As we can see in the Figure 10, our final results for the 9th and 10th of May, were satisfactory, not considering the recent market collapse on the 9th of May we had good RMSE values and for most cases a descending crypto value trend that was somehow followed in the actual market.

## 7. DEPLOYMENT AND APPLICATION TO THE BUSINESS

Because the cryptocurrency market is incredibly volatile market and the trends vary quickly over time, it is a critical step to continually "feeding" the predictive model with current data and evaluate the parameters and trends to guarantee that the model does not depreciate.

## 8. CONCLUSION

The truth is, just like every other economic variable, predicting the price of a financial asset is extremely difficult. Using technical indicators is one of the great tools for investors to have a clue about future trend shifts, but we must keep in mind that it is only focused on historical trading data. Cryptocurrency prices also rely on the perceived “inherent value” of the market and are impacted by many external parameters. Going back to a few years ago, incidents like the global economic crisis of 2007-2008 demonstrated that no statistical model, no matter how complete or precise it is, can accurately forecast future financial movements. Moreover, sentiment analysis also plays a crucial role in this subject and would be interesting work to develop in the future. As seen in Figure 11, fewer than 280 characters on twitter from Elon Musk, can cause a significant increase/decrease of the bitcoin prices.

Nonetheless, forecasting models are useful in the economic and financial sectors to help anticipate certain shocks. Moreover, by using the different indicators and observing the combinations of candlesticks, the company will be able to define patterns that will help them to better predict the short- and long-term price movements.

# 9. REFERENCES


- Elon Musk says Tesla will once again accept bitcoin. (2022). Retrieved 7 May 2022, from https://www.vox.com/recode/2021/5/18/22441831/elon-musk-bitcoin-dogecoin-crypto-prices-tesla
- Calculating the RSI in Python: 3 Ways to Predict Market Status & Price Movement - αlphαrithms. (2022). Retrieved 7 May 2022, from https://www.alpharithms.com/relative-strength-index-rsi-in-python-470209/
- Hayes, A. (2021, July 22). *Technical analyst definition*. Investopedia. Retrieved May 9, 2022, from https://www.investopedia.com/terms/t/technical-analyst.asp
- *10 trading indicators every trader should know*. IG. (n.d.). Retrieved May 9, 2022, from https://www.ig.com/en/trading-strategies/10-trading-indicators-every-trader-should-know-190604
- Lewinson, E. (2021, November 5). *Top 4 python libraries for technical analysis*. Medium. Retrieved May 9, 2022, from https://medium.com/geekculture/top-4-python-libraries-for-technical-analysis-db4f1ea87e09
- Lendave, V. (2021, December 14). *LSTM vs gru in recurrent neural network: A comparative study*. Analytics India Magazine. Retrieved May 9, 2022, from https://analyticsindiamag.com/lstm-vs-gru-in-recurrent-neural-network-a-comparative-study/

## 10. APPENDIX

**<u>Figure 1:</u>** Dataset and Cryptocurrencies Description

Dataset's

- Low: Lowest price during a day
- High: Highest price during a day
- Open: Price at the start of the day
- Close: Price at the end of the day
- Adj. Close: Closing price after adjustments for all applicable splits and dividend distributions.
- Volume: Amount of an asset or security that changes hands over the course of a day

Coin-USD

- ADA-USD: Cardano
- ATOM-USD: Cosmos
- AVAX-USD: Avalanche
- AXS-USD: Axie Infinity
- BTC-USD: Bitcoin
- ETH-USD: Ethereum
- LINK-USD: Chainlink
- LUNA1-USD: Terra
- MATIC-USD: Polygon
- SOL-USD: Solana

**<u>Figure 2:</u>** Correlation Between Cryptocurrencies

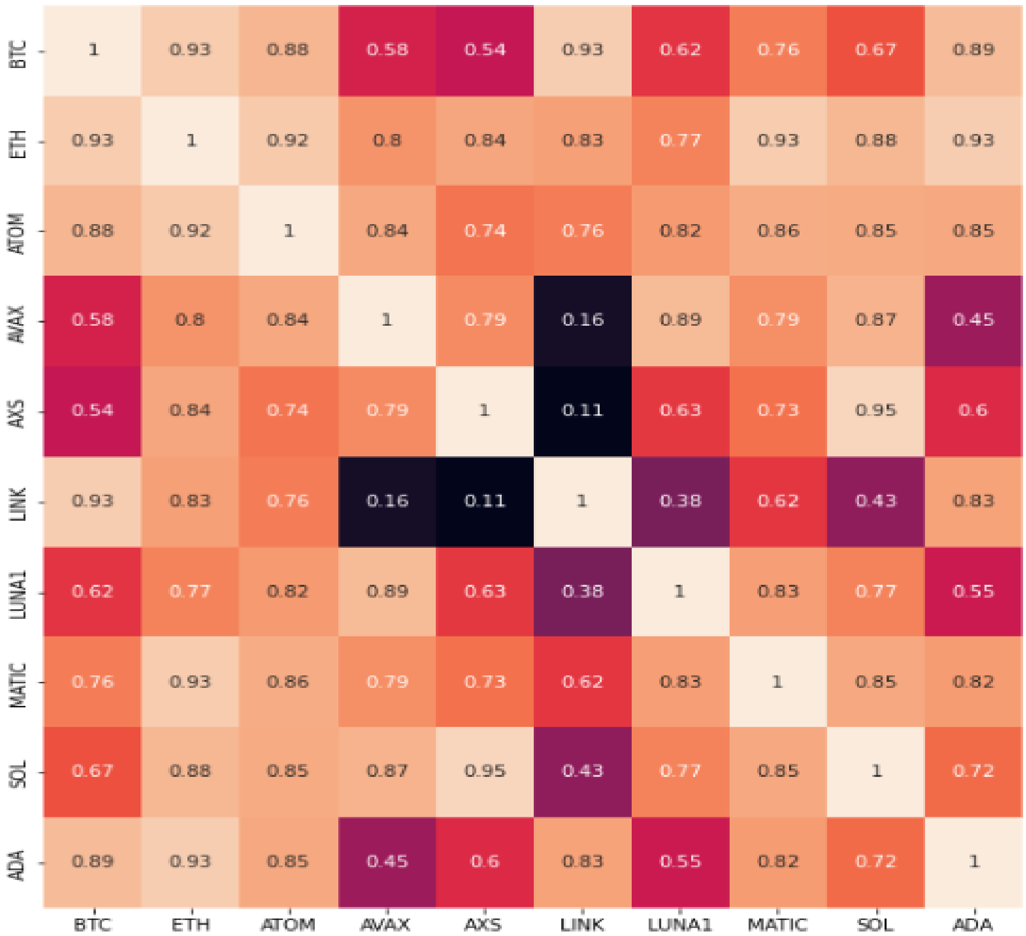

**Figure 3:** Bitcoin Candlestick with Range Slider


**Figure 4:** RSI level of Bitcoin/USD


**Figure 5:** MACD histogram of Bitcoin/USD

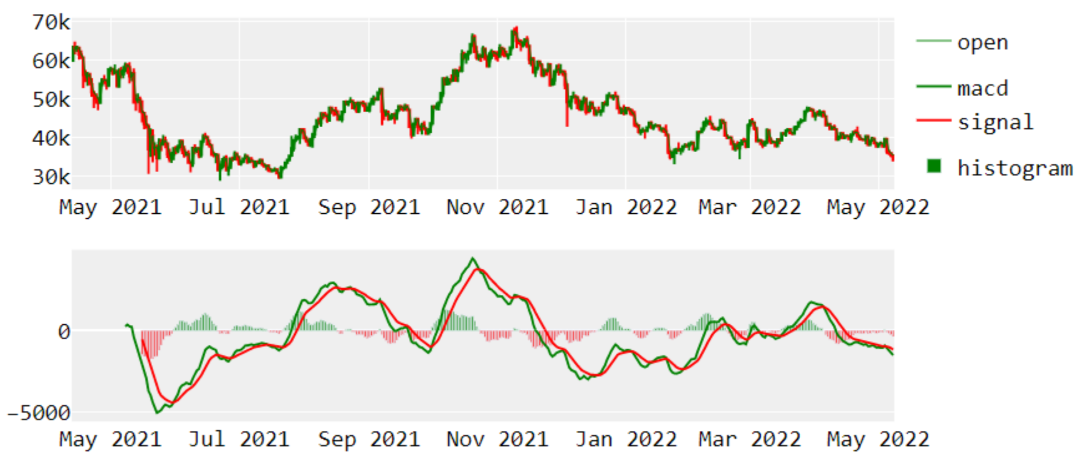

**Figure 6:** Stochastic Oscillator indicator for Bitcoin/USD

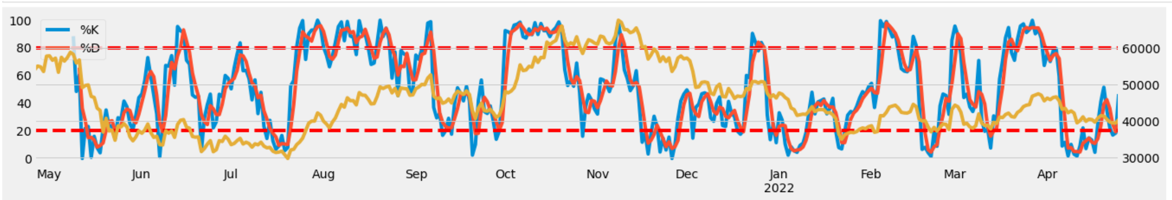


**Figure 7:** Sell and Buy signal of the Stochastic Oscillator indicator for Bitcoin/USD

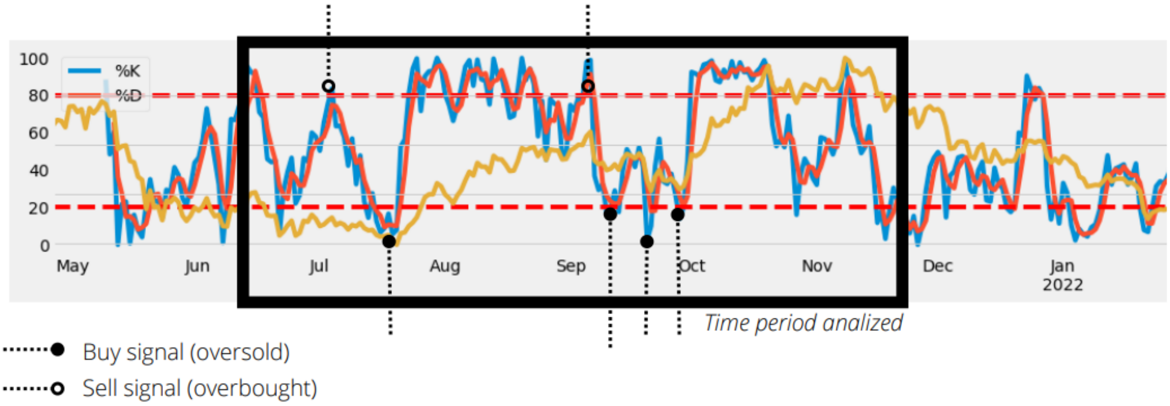

**Figure 8:** Results of applying LSTM and GRU

```
                ---- LTSM single -----

epochs -> 100 rsme -> 0.05570747535675764 4 pred -> 40203.648438
epochs -> 300 rsme -> 0.04142980167021354  pred -> 39431.230469
epochs -> 600 rsme -> 0.03534566396847367 6 pred -> 39479.562500

                   ---- GRU single ------

epochs -> 100 rsme -> 0.03658207545056939  pred -> 39840.266
epochs -> 300 rsme -> 0.031420325456808014 pred -> 39179.140625
epochs -> 600 rsme ->  0.0296375764037172  pred -> 39395.058594



                ---- LTSM multi OHLC -----

epochs -> 100 rsme -> 0.05748543415218592  pred -> 40331.957031
epochs -> 300 rsme -> 0.04289498871192336  pred -> 40100.132812
epochs -> 600 rsme -> 0.035913149444386366 pred -> 39611.554688

                 ---- GRU multi OHLC ------

epochs -> 100 rsme ->  0.036134731899946926 pred -> 39845.316406
epochs -> 300 rsme ->  0.0310757704389592   pred -> 39050.777344
epochs -> 600 rsme ->  0.029616253285979232 pred -> 39717.066406

                ---- LTSM multi Indicators  -----

epochs -> 100 rsme ->  0.05364902563393116 pred -> 39970.050781
epochs -> 300 rsme ->  0.04448282221332192 pred -> 40200.816406
epochs -> 600 rsme ->  0.04737946797162294 pred -> 40678.195312

                 ----- GRU multi Indicators ------

epochs -> 100 rsme -> 0.03629539925605059  pred -> 39637.437500
epochs -> 300 rsme -> 0.03298685263842344  pred -> 39615.128906
epochs -> 600 rsme -> 0.031664264754702645 pred -> 39320.730469


              ----- LTSM multi OHLC + Indicators ------

epochs -> 100 rsme -> 0.05553947612643242 pred -> 40901.960938
epochs -> 300 rsme -> 0.04876272858430942 pred -> 41520.339844
epochs -> 600 rsme -> 0.0481601855220894  pred -> 41407.238281

              ----- GRU multi OHLC + Indicators ------

epochs -> 100 rsme -> 0.034462566282600166 pred -> 39835.050781
epochs -> 300 rsme -> 0.03157485010723273  pred -> 39755.335938
epochs -> 600 rsme -> 0.030360100579758485 pred -> 39330.718750
```

**Figure 9:** Performance Metrics per No of Epochs

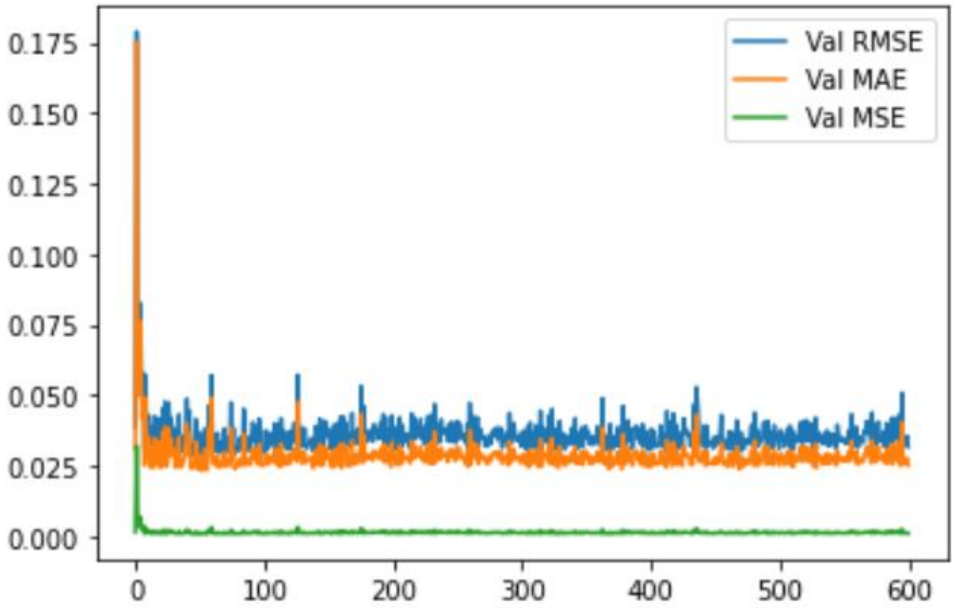


**Figure 10:** Table of the results of the prediction and RMSE evaluation

| Coin-USD | RMSE | Prediction ($) 09/05 | Prediction($) 10/05 |
|---|---|---|---|
| ADA-USD | 0.03579 | 0.81 | 0.79 |
| ATOM-USD | 0.04822 | 16.77 | 16.42 |
| AVAX-USD | 0.03204 | 55.15 | 53.34 |
| AXS-USD | 0.02055 | 28.32 | 28.42 |
| BTC-USD | 0.03600 | 35437.33 | 34815.37 |
| ETH-USD | 0.03587 | 2654.08 | 2583.53 |
| LINK-USD | 0.03778 | 12.02 | 12.24 |
| LUNA1-USD | 0.07720 | 61.97 | 57.23 |
| MATIC-USD | 0.02837 | 1.04 | 0.98 |
| SOL-USD | 0.02687 | 81.30 | 79.08 |

**Figure 11**: How Elon Musk's Tweets Have Influenced Bitcoin Prices.