Crypto Prediction using Linear Regression.

Abstract – This document shows de implementation of a linear regression using sklearn framework to determine the price of cryptocurrencies.

I. Introduction.

Cryptocurrencies have gained a lot of popularity around the world; it is a digital asset not regulated by any governmental entity or Bank. Many people started to invest on digital assets, but the prices are extremely volatile, so their price is always going up and down.

II. DATASET.

The dataset used to test this project was obtained from Kaggle [2]. This compilation of data is the historical record of Bitcoin. Usually, it is split by days, month, or years, but to have more data we are using data from 2013 to 2021. It is around 2800 records, which includes the daily price of Bitcoin. Also, a dataset of the last year was obtained from blockchain.com [3]

Comentado [PSA1]:

Date	High	Low	Open	Close	Volume	Marketcap
29/04/2013	147.488007	134	134.444	144.539993	0	1603768865
30/04/2013	146.929993	134.050003	144	139	0	1542813125
01/05/2013	139.889999	107.720001	139	116.989998	0	1298954594
02/05/2013	125.599999	92.2818985	116.379997	105.209999	0	1168517495
03/05/2013	108.127998	79.0999985	106.25	97.75	0	1085995169
04/05/2013	115	92.5	98.0999985	112.5	0	1250316563
05/05/2013	118.800003	107.142998	112.900002	115.910004	0	1288693176
06/05/2013	124.663002	106.639999	115.980003	112.300003	0	1249023060
07/05/2013	113.444	97.699997	112.25	111.5	0	1240593600
08/05/2013	115.779999	109.599999	109.599999	113.566002	0	1264049202
09/05/2013	113.459999	109.260002	113.199997	112.669998	0	1254535382
10/05/2013	122	111.551003	112.799004	117.199997	0	1305479080
11/05/2013	118.679001	113.010002	117.699997	115.242996	0	1284207489
12/05/2013	117.448998	113.434998	115.639999	115	0	1281982625
13/05/2013	118.698998	114.5	114.82	117.980003	0	1315710011
14/05/2013	119.800003	110.25	117.980003	111.5	0	1243874488
15/05/2013	115.809998	103.5	111.400002	114.220001	0	1274623813
16/05/2013	118.760002	112.199997	114.220001	118.760002	0	1325726787
17/05/2013	125.300003	116.570999	118.209999	123.014999	0	1373723882
18/05/2013	125.25	122.300003	123.5	123.498001	0	1379574546
19/05/2013	124.5	119.570999	123.210999	121.989998	0	1363204703

Figure 1: Part of the data taken for the project.

III. Approach

The purpose of this project is to determine the value of a digital asset in the next N days.

This prediction can be done with a linear regression to get an approximation between the selected data and the future values.

To do this we evaluate this we take into consideration the fields of the bitcoin price at the end of the date: Being the ‘Close’ column.

For this we take the information of the original dataset and created an independent one without N rows to fill it with the predicted values.

Actual price bitcoin fot the last 30 days

	Price	Prediction
336	57796.62	34509.454668
337	57857.50	34758.420975
338	56610.46	31347.012555
339	57213.33	32635.273472
340	53241.72	29817.044671
341	57473.23	33322.810960
342	56428.16	31812.469632
343	57380.27	33037.525019
344	58928.81	33283.111896
345	58280.73	33383.819721
346	55883.50	35920.544545
347	56750.00	31413.544922
348	49007.09	31650.912113

Figure2: Shows the comparison between the prediction vs the original price in that day. Tested on the last year dataset.

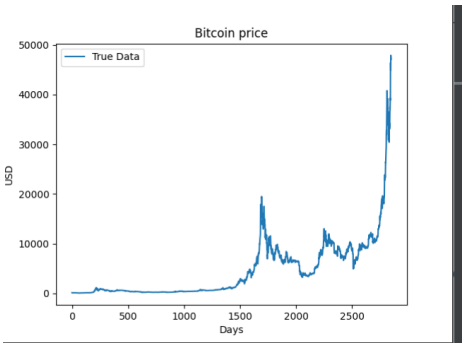

Figure3: Shows the value of the bitcoin since its release.

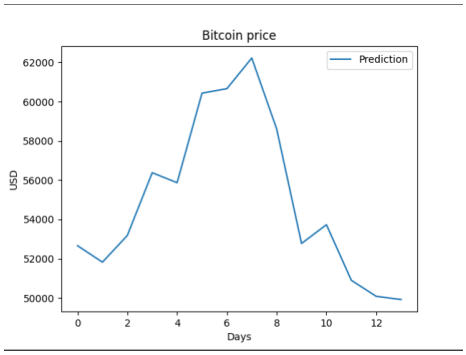

Figure4: shows the value of the bitcoin of the last 14 days.

IV Linear Regression

The data was divided in 20% for testing and 80 for training.

Linear regression was applied to calculate the value of the last N days, with the help of the sklearn linear model.

V. Results

Variance was used to measure how different were the real values from the average predicted, it returned a value of 94%.

The cross validation returned a value of 36%

VI. Conclusion

The difficulty to predict the value of a digital asset is extremely difficult, even people who are experts on this topic struggles while doing this. So this was only for learning purposes. But it may be a little bit better than trying to guess.