

Prediction of the cryptocurrency market using Azure ML

BARBAULT Cyprien
Big DATA Student
ESAIP – Engineering School
Angers, France
cbarbault.ing2022@esaip.org

GIROU Romain
Big DATA Student
ESAIP – Engineering School
Angers, France
rgirou.ing2022@esaip.org

MOTHELAY Xavier
Big DATA Student
ESAIP – Engineering School
Angers, France
xmothelay.ing2022@esaip.org

Abstract—This document is a sum up of our project mixing machine learning and cryptocurrencies. We describe the current situation of the cryptocurrency and how we decided to build our algorithm on Azure Machine Learning Studio.

Index Terms—Bitcoin, Cryptocurrency, Machine Learning, Market, Microsoft Azure

I. INTRODUCTION

For the course of Data Mining, we were given the opportunity to choose a subject to work on to apply one of the different type of AI seen in classe (Regression, Classification, NLP, Computer vision, ...). We opted for the application of regression through the prediction of the crypto-currency market.

II. WHY USING THE MACHINE LEARNING IN THE CRYPTO-CURRENCY MARKET ?

A. The world of cryptocurrencies

In 2009, the Bitcoin, future leader of the market of the cryptocurrencies was created. This money of a new type was based on the original concept of **David Chaum**. As a reminder, this is the definition of a crypto currency by the Cambridge Dictionary :

A digital currency produced by a public network, rather than any government, that uses cryptography to make sure payments are sent and received safely.
[1]

Cryptocurrencies can be “mined”¹ and offered as a reward to the “miners”. Those miners are then able to sell the fruit of their work, as a result, the crypto (short for cryptocurrency) acts like a Forex market. As miners gather more and more crypto, the quantity of currency on the market is increasing, and thus the complexity of the mining increase to avoid overflowing the market.

With the use of the **blockchain technology**, cryptocurrencies (and the Bitcoin in particular) are offering security and anonymity to its buyers and seller on every

transaction.

B. Cryptocurrencies market’s problem

Because of their lack of centralisation, cryptos-markets are highly volatiles. The Bitcoin for example have already gained more than 3000\$ in a day and lost more than 50% of its value in the same duration [2]. Crypto investor have to be really precautious with such a market, and in order to maximise their gain, the solution could be the machine learning. ML (short for Machine Learning) is the use of a big amount of data to train an algorithm in a specific task with the help of statistics.

C. Machine learning in the trading world

The bitcoin market, it’s more than 350k transactions a day ! With such a number of transactions, we have a ton of data to analyse... And people understood that, more and more people are trying to predict Bitcoin price a month, a week, a day of even a few seconds aheads. If you can predict how the currency will behave, you know when to buy and when to retire.

We now we won’t get rich with a simple regression model on Azure (or everybody would do it), but we wanted to put a foot in this field, and understand both how machine learning and markets works so we decided to see how accurately we could predict the price of various cryptos just one minute ahead

D. The Dataset

We picked a dataset based on **Gemini**’s data over the BTC/USD, ETH/USD, ZEC/USD and LTC/USD every minutes of the 2019 year.². This Dataset is made out of 1.751.122 entries ($\frac{1}{4}$ of each coin) and is based on 8 columns :

- | | |
|------------------|----------|
| • UNIX Timestamp | • Low |
| • Symbol | • Close |
| • Open | • Volume |
| • High | • Date |

¹With GPU or ASICs, not in real mines

²Bitcoin to US dollar, Ethereum to US dollar, Zcash to US dollar and Litecoin to US dollar

UNIX Timestamp and Date: The UNIX Timestamp is the elapsed time (in seconde) since the January, 1st 1970 while the date is... the date and hour of the timeframe

Symbol: The symbol of the crypto (BTC,ETH,LTC,ZEC)

Open/Close: Respectively the price of the currency at opening and closing of the market (at the beginning and the end of each minute)

High/Low: Respectively the price of the currency at its highest and lowest during the timeframe (each minute)

Volume: Number of tokens exchanged during the timeframe

E. Dataset preparation

First thing to know is that we didn't really had a dataset with the four currencies. In fact, what we really had was four datasets, each about a different crypto. So we had to concatenate the differents datasets into on big dataset. Thankfully, Azure ML is offering this kind of option in the dataset section, you can pick differents .csv files with the same header and make one big dataset of it.

The second action we took to prepare our dataset was the selection of the columns we wanted. For example, the UNIX Timestamp was worthless. We also needed to get rid of the Date column because it had differents format between the differents crypto we picked, so to avoid conflicts, we decided to remove it entirely.

For security, we also cleant the data by removing line where some fields were missing. But as the source of the datasets was **Gemini**, we didn't expected to loose a lot of it.

Then we had to normalize the Data. We picked the **MinMax** transformation method to set the value of **Open**, **Low**, **Close** and **Volume**³ in a range of 0 to 1 to speed up the training time.

Finnnaly, we splitted the data between the training set (70% of the dataset) and the test set (the 30% left). This way, we could see if the trained model is accurate or not.

F. The choice of the model

When it comes to the choice of a model for our machine learning algorithm, we initialy planned to go for the **Linear Regression Model**. But in order to be sure it was the right model for the task, we finnnaly decided to compare four differents models:

- Linear Regression
- Poisson Regression
- Machine Learning Regression
- Random Forest Regression

³These columns will be the feature for ou ML algorithm

Note that we will concentrate on the MAE (Mean Absolute Error) only for this comparison.

| Model | MAE |
|-----------------------------|-------------|
| Random Forest Regression | 0.500912 |
| Linear Regression | 1.13083 |
| Poisson Regression | 88.58514 |
| Machine Learning Regression | 3639.995401 |

TABLE I
SCORE OF THE DIFFERENTS MODELS

We had a clear winner: the **Random Forest Regression Model**

G. Results

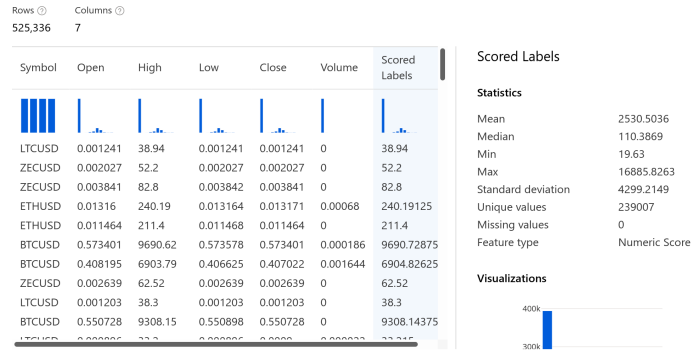


Fig. 1. Results obtained with the Random Forest Regression Algorithm

Analysis: As we can see on the figure II-G, the **High** column in the column we wanted to predict and the **Scored Labels** is the predicted result. The results are really close !

If you imagine linking an trading algorithm to this model, then you can predict the highest price of a time frame. This way, if your buying price is lower than the highest price, your trading bot could sell automatically your crypto and gain a significant margin !

III. CONCLUSION

Crypto-currencies may be the currency of the future, and we must prepare ourself for a world where the market is full of micro-transaction at really high frequency. A world where transactions could be totally anonymous and untraceable. In this world, BOT and AI would rule the market and billions of dollars would be traded each seconds. In this world there would be no place for errors... To avoid dramatics lost for some companies, AI have to be trained to predict with accuracy the fluctuations of the market.

Our goal was to predict with precision the price of the highest value of the market at every minute. Even though our algorithm seems to be really good in this cas, this prediction is on the really short term, and in the real world, it may be more useful to predict the variation at a higher scale (on a

week or a month). An improvement we could make to our project is to change the dataset for a dataset based on the evolution of the market on a daily basis or even try to predict some days in advance instead of the current minute. An other approach would be to use the sentiment analysis, for example on Twitter, to measure the confidence, this could be a good idea for a future projet.

REFERENCES

- [1] Definition of a Cryptocurrency, Cambridge Dictionary, <https://dictionary.cambridge.org/us/dictionary/english/cryptocurrency>
- [2] Charles Bovaird, Bitcoin Lost Roughly 50% Of Its Value In A Day, Mar 12 2020, <https://www.forbes.com/sites/cbovaird/2020/03/12/bitcoin-lost-roughly-50-of-its-value-in-a-day/>