Bitcoin Price Prediction Using Machine Learning, Artificial Neural Network Model and Cluster Analysis

Marianne Hanna, Re-Mi Hage

Department of Mathematics, Faculty of Natural and Applied Sciences, Notre Dame University-Louaize, Zouk Mosbeh , Lebanon

I - Introduction

Bitcoin is the most popular purely digital cryptocurrency nowadays. It started in 2009 as a peer to peer payment system, decentralized, pseudo-anonymous and secure system of money. Crypto is the new-age currency which only has an online presence and tradable nonetheless without the contribution of banks or clearinghouse. Each and every single trade is recorded in a public distributed ledger called "blockchain" presented as digital assets that serve the reward for a process known "mining" and can be exchanged for other currencies.

For the last decade, researcher have tried to estimate and predict Bitcoin using machine learning such as:

- Autoregressive Integrated Moving Average Model (Anupriya and Garg (2018)[1]),
- Decision tree, linear regression and Autoregressive Integrated Moving Average Model (Naghib Moayed and Habibi (2020)[2]),
- Simple linear, multiple regression and artificial neural network Multilayer Perceptron and Long short-term memory (Uras et al. (2020)[3]).

II - Objectives

The main objective of this study is to estimate and predict the weekly close bitcoin price by integrating other variables such as:

- Commodities like crude oil and Gold[4]
- Indexes as European, Asian and Chinese[5]
- Other cash currencies like Euro and Yuan[4,5]
- Demand/ supply variables live blockchain addresses, miner's rewards, transaction value and other variables[6].

This study reflected the short term prediction of bitcoin and compared the accuracy of different machine learning algorithm including supervised, unsupervised, and deep learning.

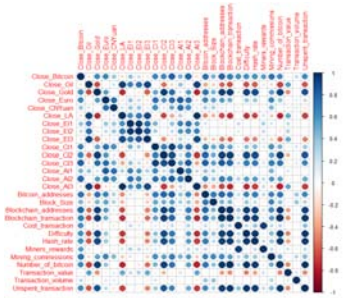

Figure 1: Correlation Plot between variables

III – Methodologies

Machine learning (ML) use algorithms and statistical models to analyze and draw inferences from patterns in data.
The following ML algorithm standalone and their combination are used to predict and estimate bitcoin before and after clustering data. Isolation forest was first used to detect in order to remove anomalies.

- Decision Tree(DT)
- Multiple Regression (REG)
- Clustering
- Time Series Autoregressive Integrated Moving Average (TSARIMA)
- Feedforward neural network (FANN)

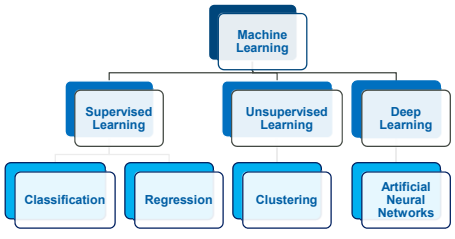

Figure 2: Machine Learning Types

IV– Algorithms

- Isolation forest is an anomaly detection algorithm. It is based on the Decision Tree algorithm.

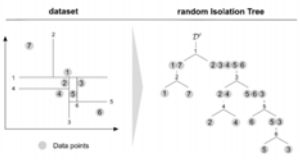

Figure 2 Isolation Forest algorithm

- DT is used to model possible consequence and map out solution to difficult situation. The result displayed on the last branch is called "node" and indicates mutually exclusive possibilities to allow the viewer understand how and why one choice may lead to the next.

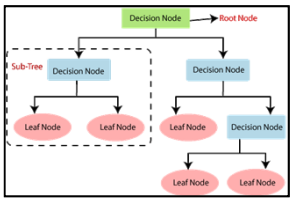

Figure 3 Decision Tree Sample

- REG is a statistical technique that model a linear relationship between several independent variables (X) and the dependent variable (Y) by which its parameters are estimated using the least square method.

$$Y_i = \sum_{k=0}^{p-1} \beta_k X_{ik} + \varepsilon_i$$

- TSARIMA used to describe certain time varying processes when data show evidence of non stationary. It consists of 3 components AR, I, MA known by ARIMA (p,d,q) where p is the lag observation, d the degree of differencing and q the number of forecast errors.

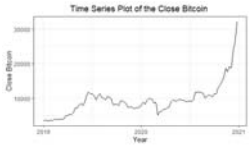

Figure 4: Time series of the weekly Close Bitcoin for the years 2019 and 2020

- FANN is a type of artificial neural network. Its information move in one direction only (input layer to the hidden layer and finally to the output) and sent from input nodes directly to output nodes. There are no feedback loops or cycles in this type of network.

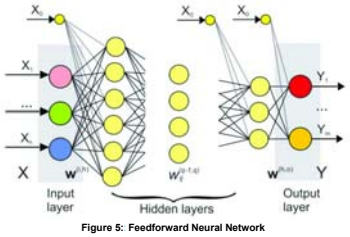

Figure 5: Feedforward Neural Network

- The goal of cluster algorithm is to group data based on as set of characteristics.

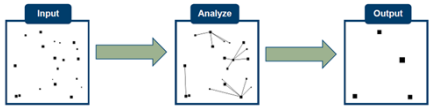

Figure 6: Cluster Analyses

V - Results

A . Before Clustering

The best model to tackle Bitcoin short term prediction is demonstrated by integrating only the variables that are reported as important by DT and the combination of regression, time series and feedforward neural network NNAR(1,1,6).

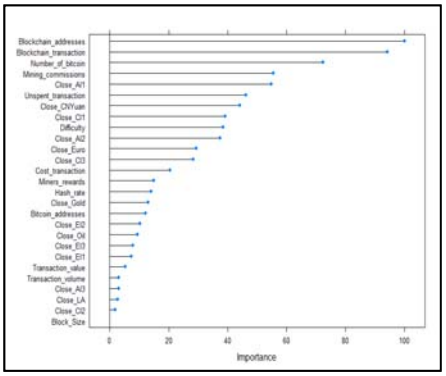

Figure 7: Important Variables through DT

ML types	MAPE
DT	6.92%
REG	10.76%
ARIMA(4,2,2)	34.77%
FANN(3)	24.70%
NNAR(1,1,6)	0.85%

Table 1: MAPE of the models

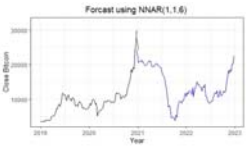

Figure 8: Close Bitcoin Forecast

B. After Cluster Analysis

In order to reduce the error, Different clustering technics such as K-Means, CLARA and PAM were tested. Pam with 4 clusters reported to be the best clustering algorithm applied on the Bitcoin data.
Clustering the data before applying NNAR have decreased significantly the MAPE and the accuracy is reported to be 99.78%.

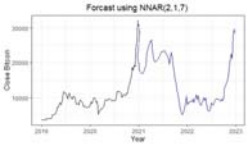

Figure 9 Close Bitcoin Forecast

VI - Conclusion

- Clustering data and combining Decision Tree, multiple linear regression, time series ARIMA and feedforward neural network increased the accuracy of estimating and predicting weekly bitcoin close price for short term.
- In the future the long term close price bitcoin prediction will be investigated by building a model with more granular variables.

References

1. Anupriya and S. Garg, "Autoregressive Integrated Moving Average Model based Prediction of Bitcoin Close Price," 2018 International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2018, pp. 473-478, https://ieeexplore.ieee.org/document/8748423
2. Naghib Moayed, A., & Habibi, R. (2020). Crypto- Currency Price Prediction with Decision Tree Based Regressions Approach. Journal of Algorithms and Computation, 52(2), 29–40. https://doi.org/10.22059/jac.2020.7911
3. Uras, N., Marchesi, L., Marchesi, M., & Tonelli, R. (2020). Forecasting Bitcoin closing price series using linear regression and neural networks models. PeerJ Computer Science, 6, e279. https://doi.org/10.7717/peerj-cs.279
4. https://www.investing.com/
5. https://www.quandl.com/
6. https://www.blockchain.com/

Contact Information

- Dr. Re-Mi Hage, Associate Professor, Academic Advisor for the BS and MS Actuarial Sciences Program
  T: +961 9 218 950 Ext: 2106 F: +961 09 218 771
  M: +961 70 885956 e-mail: rhage@ndu.edu.lb
- Marianne Hanna, MS Actuarial Sciences
  M: +961 76 101297 e-mail: mghanna01@ndu.edu.lb