New predictive opportunities for decentralized blockchain assets

CallBot Team
Polytechnic University of Turin
Vienna University of Technology
Technical University of Munich

***Abstract*—Blockchains have been revolutionary by allowing anyone to own and transfer assets across an open financial network without the need for a trusted third party. Cryptocurrencies are digital coins not issued by any government or legal entity; they only use cryptography and suitable logical systems to regulate their supply, control trading operations and avoid frauds. The transactions are recorded digitally in a blockchain as an accounting system. The fact that the blockchains data are public allows investors to have much more knowledge about the assets they are investing in. In this work we present the main characteristic of a novel mathematical model which aims to identify the trend of cryptocurrencies, based on the internet and blockchain data. The algorithms present over 70 input parameters, which include Google trends, various sentiment analyses, Twitter, Instagram and Reddit tendencies. The main idea is to detect secure investment opportunities with the application of three completely different predictive algorithms. A signal comes when the three algorithms predict the same trend of a cryptocurrency. The three algorithms used, have different mathematical roots: Statistical, deep learning and autoregressive. The results of our work show the suitability and the effectiveness of using three different algorithms to predict the cryptocurrencies price trends.**

***Index Terms*—Predective algorithms, Blockchain, Machine Learning, Non-Linear models**

I. Introduction

The combined success of the open-source ecosystem, decentralized file sharing and public cryptocurrencies has inspired several applications of the blockchain technologies. Cryptocurrencies are utilized worldwide for computerized investment. There are major differences between the stocks and the cryptocurrencies. The cryptocurrencies are functionalities are open source and investors have much more knowledge about the assets they are investing in in comparison the the stock market. There are various algorithms using machine learning which are utilized for price prediction on stocks value. The limit consists in hidden information (e.g. the number of transaction or the number of holders) which reduces the quality of the stocks price predictions.

The information about the blockchains is mostly open source and can be analysed in real-time. For this reason, blockchains present new predictive opportunities for the decentralised assets. Thus, prediction algorithms gain importance and potential for the cryptocurrencies. The main idea of this work is to predict the short-term trend of volatile cryptocurrencies based on several public parameters. The novel predective algorithm takes into account over 70 input parameters, which include Google trends, various sentiment analyses, Twitter, Instagram and Reddit tendencies. Recent studies in big data analytics and natural language processing develop automatic techniques in analyzing sentiment in the social media information. In addition, the growing user base of social media and the high volume of posts also provide valuable sentiment information to predict the price fluctuation of the cryptocurrency. This research is directed to predicting the volatile price movement of cryptocurrencies by analyzing the public market data and the sentiment in social media and finding the correlation between them. For this purpose three completely different predictive algorithms are applied and an investment opportunity is identified when the three algorithms predict the same trend of a cryptocurrency.

This document is organised as follows. Section II presents the driving factors that leaded to new predictive opportunities for decentralized presenting the input parameters used in the developed algorithms. Section III illustrates the algorithms used to predict the short-term trend of volatile cryptocurrencies and section IV shows the experimental results achieved. Finally, section V concludes the paper and elaborates on possible directions for future research and improvement.

II. Driving factors

In this paper, we aim to develop an algorithm to predict the prices of several cryptocurrencies accurately taking into consideration various parameters. The fact that the blockchains data are public allows investors to have much more knowledge about the assets they are investing in. Thus, applied predictive algorithm have wide information about the investigated assets. The algorithms present over 70 input parameters categorized in the following families.

- Market data
- Google Trends data
- Blockchain data
- Twitter data
- Instagram data
- Telegram data
- Discord data
- Reddit data

Market data, like volume traded, price volatility, market capitalization etc. are extrapolated from the major cryptocurrency exchanges. Based on this data several market indicators are calculated, such as RSI and EMA and used as input for the

predictive algorithms. Google Trends data give insights about the interest of people in searching a certain cryptocurrency on google. The blockchain data include the number of transactions, the number of holders and the information about the blocks. Twitter and instagram data are significant in terms of number of hashtags, number of followers and sentiment analysis. This data are extrapolated and taken into account by the predictive algorithm. Finally, various sentiment analyses are performed on selected telegram chats, discord communities and reddit groups. In the following chapter the algorithms used to predict the short-term trend of volatile cryptocurrencies with the above mentioned parameters are illustrated.

III. Method

The main idea is to detect secure investment opportunities with the application of three completely different predictive algorithms. The three algorithms used, have different mathematical roots: Statistical, deep learning and autoregressive.

The statistical method used is a nonlinear regression. In statistics, nonlinear regression is a form of regression analysis in which observational data are modeled by a function which is a nonlinear combination of the model parameters and depends on one or more independent variables. The data are fitted by a method of successive approximations as shown below.

$$y \sim f(x, \beta) \tag{1}$$

$$f(x, \beta) = \frac{\beta_1 x}{\beta_2 + x} \tag{2}$$

In eq. 1 relates the vector of independent variables (or input parameters) x, and its associated observed dependent variables y (in other words the historical prices). The function f is nonlinear in the components of the vector of parameters β, but otherwise arbitrary. In general, there is no closed-form expression for the best-fitting parameters, as there is in linear regression. Usually numerical optimization algorithms are applied to determine the best-fitting parameters. Again in contrast to linear regression, there may be many local minima of the function to be optimized and even the global minimum may produce a biased estimate. In practice, estimated values of the parameters are used, in conjunction with the optimization algorithm, to attempt to find the global minimum of a sum of squares.

Furthermore, a deep learning technique is applied. An artificial neural network (ANN) is developed and trained with the input parameters presented in section II. An ANN is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal to other neurons, as shown in figure 1 .

An artificial neuron receives a signal then processes it and can signal neurons connected to it. The ”signal” at a connection is a real number, and the output of each neuron is computed by some non-linear function of the sum of its inputs. Neural networks learn (or are trained) by processing examples, each of which contains a known ”input” and ”result,” forming probability-weighted associations between the two, which are

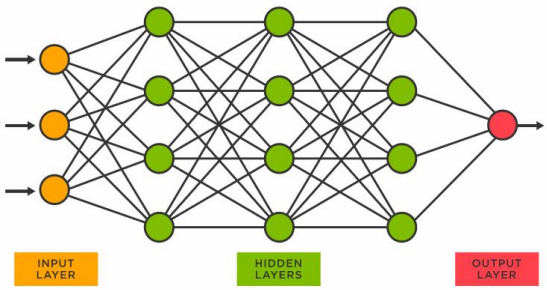

Fig. 1. Architecture of an artificial neural network

stored within the data structure of the net itself. The training of a neural network from a given example is usually conducted by determining the difference between the processed output of the network (often a prediction) and a target output. This difference is the error. The network then adjusts its weighted associations according to a learning rule and using this error value. The ANN applied in our work is composed by 4 layers and the loss function is based on the mean squared error. The batch size is of 50 example and the epochs are 250 in the learning process. These parameters are shown to be the most suitable for the short-term prediction.

Finally, an autoregressive prediction model is used. More specifically the autoregressive integrated moving average (ARIMA). In statistics and econometrics, and in particular in time series analysis, an ARIMA model is a generalization of an autoregressive moving average (ARMA) model. Both of these models are fitted to time series data either to better understand the data or to predict future points in the series (forecasting). The AR part of ARIMA indicates that the evolving variable of interest is regressed on its own lagged values. The MA part indicates that the regression error is actually a linear combination of error terms whose values occurred contemporaneously and at various times in the past. The I (for ”integrated”) indicates that the data values have been replaced with the difference between their values and the previous values (and this differencing process may have been performed more than once). The purpose of each of these features is to make the model fit the data as well as possible.

The following section shows the experimental results achieved with the use of the above mentioned algorithms feeded with the input parameters presented in section II.

IV. Experimental results

The first experimental results of the novel predictive approach have been collected between the 24.03.2022 and the 26.03.2022. The model identify an investment opportunity when the above mentioned algorithms predict the same trend of a cryptocurrency. In the investigated period three investment opportunities have been detected. The investment opportunities are resumed in figure 2, 3 and 4, respectively.

The algorithms are set up to detect profit opportunities of 4+%. The learning process of the algorithm considers a signal wrong if the price drops by 3% before increasing by