

Crypto-Asset Trading Analysis

Harsh Wasan¹, Rajdeep Khandelwal², Hemita Haldankar³, Prof. Swati Sharma⁴

^{1,2,3}Student, IT Department, Vidyalkar Institute of Technology, Mumbai, India

⁴Professor, IT Department, Vidyalkar Institute of Technology, Mumbai, India

Abstract: Initial coin offerings (ICOs) are a new method of raising capital for early stage ventures, an alternative to more traditional sources of start-up funding. Similar to the stock market the Crypto currency market has experienced a growth in various investing options for the investors. On the block chain one can trade crypto currencies with actual real-life assets and one can also invest in various fundraisers of startups for ICO's which can be traded for monetary value in the startup's environment. Investing in ICOs can be very profitable. Picking good token for investment requires careful research. We are building a software which helps to predict the prices of ICO's so that the investors can decide whether to invest in ICO or not. Since this is an emerging field there are no major applications or websites which provide consulting services for the crypto assets. We can access past data and use AI to forecast and predict future trends and advise investors accordingly

Keywords: Crypto-currency, ICO, AI and ML, Block chain, Funds.

1. Introduction

An initial coin offering (ICO) or initial currency offering may be a sort of funding using crypto currencies. Mostly the method is completed by crowd funding but private ICOs are getting more common. In ICO, a quantity of crypto currency is sold in the form of "tokens" ("coins") to speculators or investors, in exchange for legal tender or other crypto currencies such as Bitcoin or Ethereum. The tokens sold are promoted as future functional units of currency if or when the ICO's funding goal is met and therefore the project launches. In ICO, a block chain-based issuer sells cryptographically secured digital assets, usually called tokens. Explosive growth in the ICO market has attracted interest from entrepreneurs, investors, and regulators. According to one estimate, between January 2014 and December 2018 ICOs raised over \$28 billion, and at least 15 individual ICOs to date have taken in more than \$100 million. At the same time, the market has become notorious for scams, jokes, and frauds. This paper asks which venture and ICO process characteristics predict real and financial success for ICO issuers, focusing on whether the market exhibits dynamics consistent with existing theoretical literature about entrepreneurial finance and, more recently, about ICOs. ICOs can provide more security, liquidity and transparency than conventional financing instruments. These features potentially mitigate costs of asymmetric information and agency problems that have long deterred arms-length retail

investment in early stage private ventures (Hall and Lerner 2010).

There are three types of digital assets.

1. The first is a general-purpose medium of exchange and store of value crypto currency, such as Bit coin; these are often termed coins.
2. The second is a security token, which represents a conventional security that is recorded and exchanged on a block chain to reduce transaction costs and create a record of ownership.
3. The third is a utility token, which gives its holder consumptive rights to access a product or service.

Utility tokens comprise the largest and most well-regarded ICOs and are the primary focus of who want to invest in crypto currency. For example, Ether (the token of the Ethereum block chain) is a utility token, but its widespread circulation has led it to become also a store of value. Utility token ICOs somewhat resemble crowd funding pre-sales of products on platforms such as Kick starter. Perhaps a closer analogy is selling tradable ownership rights to stadium seats before a sports or entertainment venue is built, a practice that goes back to the 19th century. While utility tokens can be simple "corporate coupons" that give the holder the right to an issuer's product or service, the most well-known ICOs employ them as the means of payment in a new marketplace. In this case, we can extend the analogy to suppose that the unbuilt stadium's games were to be played (or at least watched) by people in the grandstands.

Proponents argue that block chains with native tokens permit disintermediation of Internet marketplaces such as Uber or Facebook. In these traditional models, the developing firm's control over the platform enables it to extract a large share of the platform's surplus, and this control also raises concerns over the developer's use of transacting party data. The token's value is often expected to increase with the value of the decentralized network. This correspondence enables three features, though not every ICO makes use of all three. First, the token can reward the network creators without giving them control after the network has launched.

Second, token buyers can fund the platform's development, speculating on the long-term value of the service it will provide in the future.

Third, like concert tickets, food stamps, or stock certificates, the token's value is tied to access to a future good or service, creating customer commitment.

In a study a sample of 100 geographically dispersed ICOs, for which they collected data on a wide range of characteristics, such as whether the token has utility value, previous VC financing, and founder professional backgrounds. The practice of raising capital from prospective customers by selling ownership rights for future seats in an not built arena dates back at least to Royal Albert Hall in London in the 1860s. Others trace the practice to the time of the Reformation or even earlier, when European church construction began to be financed by the advance sales of pews that were owned in perpetuity by their sponsors and could be re-sold for profit. For utility tokens to have value, the issuer must commit to a cap on the total supply, and this is easily done in a sports arena or church where adding new seats are physically difficult. Smart contracts can impose these limits for ICO tokens. They analyzed that ICO possess characteristics for a subsample of some successful offerings that subsequently traded on secondary market exchanges for at least 10 days. Following From the perspective of an early stage investor, liquidity represents a central benefit of ICOs relative to conventional financing instruments. Liquidity also reflects market depth and interest in a token. Sockin and Xiong (2018) show that token trading enables information aggregation from potential customers about demand for a platform's service, and they conclude that an individual's decision to join a token-based platform depends positively on the token's trading volume. However, the liquidity of ICOs may have a dark side if issuers' ability to cash out quickly undercuts their incentives to build successful businesses, or if investors do not have incentives to monitor intensively. Consistent with the previous analysis of real outcome predictors, they found that liquidity and trading volume are higher for tokens that offer voluntary disclosure in a white paper, credibly commit to the project through insider vesting restrictions, and signal quality via prior VC investment and past entrepreneurial success of the CEO. For example, success is associated with token sales that use dynamic pricing mechanisms, that promote transparency and crowd source development by publicly posting source code on Github, and that have large Telegram user groups. In contrast, asset management and other crypto financial services are if anything negatively associated with success. These results shed light on where the market has perceived opportunities for value creation. So using some of the afore mentioned factors we analyze the current live ico's to predict their costs in order to advice the investors whether they should invest in a particular ico or not. We create a web- app which acquires data from an api and then the data is analyzed

using ml and then the forecasted costs are shown to the investors to act as an aid.

2. Literature Survey

Factors such as-target proceeds, fraction of total token supply sold, pricing mechanism, distribution method, lock-ups and set-asides, token rights, and selection of secondary market exchange [1] affect the cost of the ICO's. A significant predictor of survival and employment is whether a token has apparent utility value, which has strong relevance for current policy debates over whether ICO tokens are investment securities in disguise, or whether they represent an innovation that enables a new venture to raise funds in a way that promotes future product adoption and loyalty, while also offering liquidity. Additional factors associated with ICO success reflect longstanding theories in corporate finance about the importance of reducing information asymmetry and the use of bonding and certification strategies to reduce agency costs. First, they examined factors that predict listing these largely parallel those that predict survival and employment. Second, noted that listing is itself an interesting characteristic, with a strong connection to token liquidity, they instrument for listing to assess its effect on employment. Specifically, they used price changes in the Ethereum Classic (ETC) token around the time of an ICO, focusing on periods when Bit coin prices are high

They also assessed which characteristics exhibit significant associations with secondary market liquidity and trading volume.

Piotr Płoński in his medium post displays how to use ML to predict the cost of the ICO's he says that the following factors [2] affect the ICO's cost.

- ICO Price
- Number of ICO tokens (number of tokens available for sale in ICO)
- Total Supply Ratio (number of tokens in ICO divided by total token number)
- ICO Market Cap
- Information if a prototype is available
- Number of points for team scored by Ian(analyst)
- Number of points for advisors scored by Ian(analyst)
- Number of points for ICO idea
- The output of the model will be Max_CMC_x which is defined as:

$$\text{Max_CMC_x} = \text{Max_CMC} / \text{ICO_Price}$$

From a banker's perspective, the problem with crypto currencies [3] is not having insight into cashflows, which is needed if one wants to both take deposits and lend - cashflow performance equals credit worthiness. In this sense, banks

have a competitive advantage over crowdfunding and other decentralised innovations, including Bitcoin (Dimon, 2016)

The first step in collecting data about each project is to collect information from the most used Internet sources as icobench, TokenData or similar. In this step we look for general characteristics such as the name, the token symbol, start and end dates of the crowdfunding [4], the country of origin, financial data such as the total number of issued token, the initial price of the token, the platform used, data on the team proposing the ICO, data on the advisory board, data on the availability of the website, availability of white paper and social channels. Some of these data, such as short and long description, and milestones are textual descriptions. Others are categorical variables, such as the country, the platform, the category (which can assume many values), and variables related to the team members (name, role, group). The remaining variables are numeric, with different degrees of discretization. Unfortunately, not all ICOs record all variables, so there are several missing data. The ICO web databases that we use are fully checked in order to minimize the missing values of one of the platforms, therefore we validate the information checking for the details on the website and on the white paper. As a result, the complete set of reliable information comes from the matching between the website and the white paper

The independent variable for this study is the closing price of Bitcoin in US Dollars taken from the [5] Coindesk Bitcoin Price Index. Rather than focusing on one specific exchange this price index takes the average prices from five major Bitcoin exchanges; Bitstamp, Bitfinex, Coinbase, OkCoin and itBit. If one were to implement trades based on the signals it would be beneficial to focus on one exchange.

The standard approach for asset value predictions is based on market analysis with an LSTM(Long short-term memory) [6] neural network. Block chain technologies, however, give us access to vast amounts of public data, like the executed transactions and therefore the account balance distribution. We explore whether analysing this data with modern Deep Learning techniques leads to higher accuracies than the quality approach

Initial ICOs [7] showed that one among the most important problem of this fundraising method was the danger of fraud and absence of trust to founding teams because the whole process was administered entirely online. Quickly, some individuals offered escrow services by acting as a trusted third party that collected, held, disbursed funds according to predetermined rules, usually by reaching some milestones announced by the ICO teams.

Linear Regression is a machine learning algorithm based on supervised learning [8]. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables, they are considering and the number of independent variables being used.

Random forests or random decision forests are an ensemble learning method for classification, regression [9] and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of over fitting to their training set

The following are the advantages of Random Forest algorithm over most of the other algorithms

- It overcomes the problem of over fitting by averaging or combining the results of different decision trees.
- Random forests work well for a large range of data items than a single decision tree does.
- Random forest has less variance than single decision tree.
- Random forests are very flexible and possess very high accuracy.
- Scaling of data does not require in random forest algorithm. It maintains good accuracy even after providing data without scaling.
- A Random Forest algorithm maintains good accuracy even a large proportion of the data is missing.

The K-fold Cross Validation (KCV) technique is one of the most used approaches by practitioners for model selection and error estimation of classifiers. The KCV consists in splitting a dataset into k subsets; then, iteratively, some of them are used to learn the model, while the others are exploited to assess its performance. However, in spite of the KCV [10] success, only practical rule-of-thumb methods exist to choose the number and the cardinality of the subsets. We propose here an approach, which allows to tune the number of the subsets of the KCV in a data-dependent way, so to obtain a reliable, tight and rigorous estimation of the probability of misclassification of the chosen model.

3. Proposed System

Block Diagram

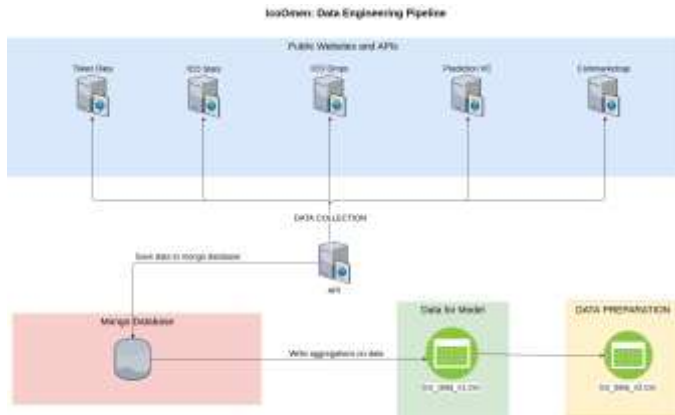


Fig. 3.1 Overall system block diagram

```

{
  name: "Augur",
  image: "https://icowatchlist.com/logos/augur.png",
  description: "Augur is a decentralized prediction market",
  website_link: "https://api.icowatchlist.com/public/v1/wc1/augur",
  icoatchlist_url: "https://icowatchlist.com/ico/augur",
  start_time: "2015-05-17 00:00:00",
  end_time: "2015-10-01 00:00:00",
  timezone: "UTC+0",
  coin_symbol: "REP",
  price_usd: "21.96",
  all_time_roi: "3,680.44%",
},
{
  name: "link",
  image: "https://icowatchlist.com/logos/link.png",
  description: "It is a cryptocurrency and decentralized application platform",
  website_link: "https://api.icowatchlist.com/public/v1/wc1/link",
  icoatchlist_url: "https://icowatchlist.com/ico/link",
  start_time: "2016-02-22 00:00:00",
  end_time: "2016-03-21 00:00:00",
  timezone: "UTC+0",
  coin_symbol: "LSK",
  price_usd: "2.87",
  all_time_roi: "2,672.96%",
},

```

Fig. 3.2 – Data received from api

In order to mitigate the risk involved in investing in ICO's we will use machine learning on the data collected by us by call an API which returns data about the current ICO's and stored in the database. Once the model is trained we apply the trained model to the real time data to predict the variance in the cost of the individual ICO's. In order to mitigate the risk involved in investing in ICO's we will use machine learning on the data collected by us by call an API which returns data about the current ICO's and stored in the database. Once the model is trained we apply the trained model to the real time data to predict the variance in the cost of the individual ICO's.

In this way the investors can easily see the real time data of the ICO and also see what the ML model predicts about its future cost. We also display the data in a graphical format which makes it easier for the investors to decide for themselves whether to invest in the ICO or not.

In this way the investors can easily see the real time data of the ICO and also see what the ML model predicts about its future cost. which makes it easier for the investors to decide for themselves whether to invest in the ICO or not.

4. Hardware and Software Requirements

Software requirements:

- Node.js
- .Net(if used as an web service)
- SQL/mongo dB database

Hardware requirements:

- Laptop (to access the web-app)

5. Methodology

Step 1:

Creating Input data:

- We are using past data of the ICOs. Data is fetched by using API:- <https://api.nomics.com/v1/currencies/ticker>
- The dataset comprises of ICO_Name, Ian_ICO_Grade, ICO_price, Number_of_ICO_Tokens,ICO_Market_Cap, Total_Supply_Ratio, Max_CMC_x indicators.
- The ICO price is the most important feature for predicting ICO returns.
- Any Analyst's grade is the second most important feature, the analyst's grade is correlated with the project's quality. The higher the grade is, the better project is, and the higher chance for the price growth.
- There are also number of tokens sold in ICO, ICO market cap, and total supply it is similar situation to that of the ICO's price.
- Data files are stored in .csv format

Step 2:

Training the Data

Next step is to train the gathered data set by using following machine learning algorithms

1. Regression (Random decision forest)
2. K-fold cross validation.

Random forest

Random forest is a supervised learning algorithm which is used for both classification as well as regression. But however, it is mainly used for classification problems [9]. As we know that a forest is made up of trees and more trees means more robust forest. Similarly, random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result.

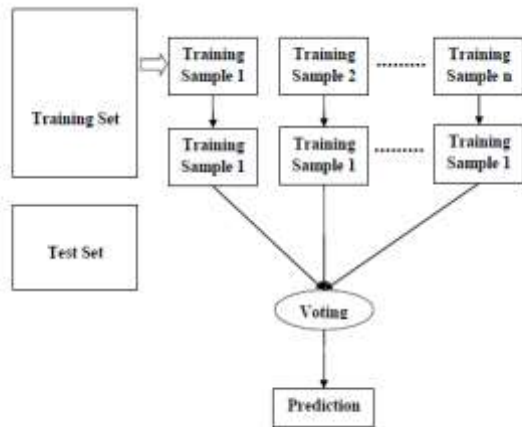


Fig 5.1

Working of Random Forest Algorithm

- **Step 1** – First, start with the selection of random samples from a given dataset.
- **Step 2** – Next, this algorithm will construct a decision tree for every sample. Then it will get the prediction result from every decision tree.
- **Step 3** – In this step, voting will be performed for every predicted result.
- **Step 4** – At last, select the most voted prediction result as the final prediction result.

k-Fold Cross-Validation

Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample.

The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into. As such, the procedure is often called k-fold cross-validation. When a specific value for k is chosen, it may be used in place of k in the reference to the model, such as k=10 becoming 10-fold cross-validation.

Cross-validation is primarily used in applied machine learning to estimate the skill of a machine learning model on unseen data. That is, to use a limited sample in order to estimate how the model is expected to perform in general when used to make predictions on data not used during the training of the model.

The purpose of cross validation is not to help select a particular instance of the classifier (or decision tree, or whatever automatic learning application) but rather to qualify the model, i.e. to provide metrics such as the average error ratio, the deviation relative to this average etc. which can be useful in asserting the level of precision one can expect from the application. One of the things cross validation can help assert is whether the training data is big enough.

After training the models and finding the model which has the most accuracy we will then forecast the cost of the individual ICO and display it to the user

Checking whether the cost is going to rise in the future or not one can make a simple statement that whether one must invest in the ICO or not.

Step 3:

Creating User Interface:

We are using HTML, CSS, BOOTSTRAP, JavaScript and PHP for designing front-end of website.

Through this user can interact with software and can see the required predicted results.

Our front-end interfaces are going to be user-friendly for both new and professional traders. Typical functions will be presented with intuitive visual feedback that is common to other prevailing platforms. Our platform is modular, lightweight and extendable. This ensures that we offer top-notch services for all our users while retaining our usability and productivity.

6. Result



Fig. 6.1 User Interface

#	Name	Market Cap	Price
1	Bitcoin	\$189715265180	\$10415.27442508
2	Ethereum	\$30312850121	\$276.36784037
3	Tether	\$4275373763	\$0.99684206
4	ChainLink	\$1494087590	\$4.10002293
5	Maker	\$547345931	\$629.34047662
6	Lisk	\$254043413	\$1.83979924

Fig. 6.2 List of ICO

References

[1] Sabrina T. Howell, Marina Niessner ,David Yermack
"INITIAL COIN OFFERINGS: FINANCING GROWTH WITH
CRYPTOCURRENCY TOKEN SALES"

<https://www.nber.org/papers/w24774.pdf>

[2] <https://medium.com/@MLJARofficial/predicting-ico-returns-with-machine-learning-af6108ab9e39>

[3]

Initial Coin Offering (ICOs) Risk, Value and Cost in Blockchain
Trustless crypto-markets Percy Venegas

https://sci-hub.ren/https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3012238

[4] <http://economieaweb.unipv.it/wp-content/uploads/2018/01/DEMWP0164.pdf>

Paola Cerchiello (Università di Pavia) Anca Mirela Toma
(Università di Pavia), July 2018

[5]<https://arxiv.org/pdf/1810.06696.pdf> Zvezdin
Besarabov, Todor Kolev, 15 Oct 2018

[6] [http://trap.ncirl.ie/2496/McNally, Sean, \(2016\) .](http://trap.ncirl.ie/2496/McNally, Sean, (2016) .)

[7]<https://sciHub.ren/https://jai.ijournals.com/content/21/4/13.abstract> Dmitri Boreiko, Gioia Vidusso, October 2018

[8]<https://www.geeksforgeeks.org/ml-linear-regression/>

[9]https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_random_forest.htm

[10] "The 'K' in K-fold Cross Validation"

Davide Anguita, Luca Ghelardoni, Alessandro Ghio, Luca Oneto and Sandro Ridella 2012

<https://www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2012-62.pdf>

[11]

<https://towardsdatascience.com/icoomen-using-machine-learning-to-predict-ico-prices-29fa4cec6d86>