

Initial Coin Offerings: risk or opportunity?

Anca M. Toma^{1*}, Paola Cerchiello¹

¹University of Pavia, Italy

Submitted to Journal:
Frontiers in Artificial Intelligence

Specialty Section:
Artificial Intelligence in Finance

Article type:
Original Research Article

Manuscript ID:
499201

Received on:
19 Sep 2019

Revised on:
27 Feb 2020

Frontiers website link:
www.frontiersin.org

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

The paper is the product of full collaboration among the authors, however, PC has inspired the idea, the methodology and wrote section 2, 4 and 7, AMT collected the data, run the analysis and wrote section 1, 3, 5 and 6.

Keywords

ICO, Cryptocurrencies, fundraising, Classification model, Text analysis & mining, SCAM

Abstract

Word count: 189

Initial coin offerings (ICOs) are one of the several by-products of the cryptocurrencies world. Start-ups and existing businesses are turning to alternative sources of capital as opposed to classical channels like banks or venture capitalists. They can offer the inner value of their business by selling 'tokens', i.e. units of the chosen cryptocurrency, like a regular firm would do by means of an Initial Public Offering (IPO).

The investors, of course, hope for an increase in the value of the token in the short term, provided a solid and valid business idea typically described by the ICO issuers in a white paper.

However, fraudulent activities perpetrated by unscrupulous actors are frequent and it would be crucial to highlight in advance clear signs of illegal money raising.

In this paper, we employ statistical approaches to detect what characteristics of ICOs are significantly related to fraudulent behaviour.

We leverage a number of different variables like: entrepreneurial skills, Telegram chats and relative sentiment for each ICO, type of business, issuing country, team characteristics. Through logistic regression, multinomial logistic regression and text analysis, we are able to shed light on the riskiest ICOs.

Contribution to the field

Initial coin offerings (ICOs) are one of the several by-products of the cryptocurrencies world. Start-ups and existing businesses are turning to alternative sources of capital as opposed to classical channels like banks or venture capitalists. They can offer the inner value of their business by selling 'tokens', i.e. units of the chosen cryptocurrency, like a regular firm would do by means of an IPO. The investors, of course, hope for an increase in the value of the token in the short term, provided a solid and valid business idea typically described by the ICO issuers in a white paper. However, fraudulent activities perpetrated by unscrupulous actors are frequent and it would be crucial to highlight in advance clear signs of illegal money raising. In this paper, we employ statistical approaches to detect what characteristics of ICOs are significantly related to fraudulent behaviour. We leverage a number of different variables like entrepreneurial skills, Telegram chats and relative sentiment for each ICO, type of business, issuing country, team characteristics. Through logistic regression, multinomial logistic regression and text analysis, we are able to shed light on the riskiest ICOs.

Funding statement

This research has received funding from the European Union's Horizon 2020 research and innovation program "FIN-TECH: A Financial supervision and Technology compliance training program" under the grant agreement No 825215 (Topic: ICT-35-2018, Type of action: CSA)

Ethics statements

Studies involving animal subjects

Generated Statement: No animal studies are presented in this manuscript.

Studies involving human subjects

Generated Statement: No human studies are presented in this manuscript.

Inclusion of identifiable human data

Generated Statement: No potentially identifiable human images or data is presented in this study.

In review

Data availability statement

Generated Statement: The datasets generated for this study are available on request to the corresponding author.

In review

Initial Coin Offerings: risk or opportunity?

Anca Mirela Toma¹, Paola Cerchiello¹

¹ *University of Pavia, Department of Economics and Management Science, Pavia, Italy*

Correspondence*:

Anca Mirela Toma

ancamirela.toma01@universitadipavia.it

2 ABSTRACT

3

4 Initial coin offerings (ICOs) are one of the several by-products in the world of the cryptocurrencies.
 5 Start-ups and existing businesses are turning to alternative sources of capital as opposed to
 6 classical channels like banks or venture capitalists. They can offer the inner value of their business
 7 by selling 'tokens', i.e. units of the chosen cryptocurrency, like a regular firm would do by means
 8 of an IPO. The investors, of course, hope for an increase in the value of the token in the short
 9 term, provided a solid and valid business idea typically described by the ICO issuers in a white
 10 paper. However, fraudulent activities perpetrated by unscrupulous actors are frequent and it
 11 would be crucial to highlight in advance clear signs of illegal money raising. In this paper, we
 12 employ statistical approaches to detect what characteristics of ICOs are significantly related to
 13 fraudulent behaviour.

14 We leverage a number of different variables like: entrepreneurial skills, Telegram chats and
 15 relative sentiment for each ICO, type of business, issuing country, team characteristics. Through
 16 logistic regression, multinomial logistic regression and text analysis, we are able to shed light on
 17 the riskiest ICOs.

1 KEYWORDS:

18 ICOs, cryptocurrencies, fundraising, classification models, text analysis, scam

2 INTRODUCTION

19 Initial Coin Offerings (ICOs) can be considered as an innovative way of obtaining funding, promoted by entrepreneurial companies
 20 that base their business projects on a new technology known as blockchain. Up to the present date, more than 1700 cryptocurrencies
 21 have been created but not all of them are successful or characterized by a significant impact. ICOs issue "tokens", i.e. the unit of
 22 a chosen cryptocurrency, in exchange of a fiat cryptocurrency, in order to participate in the crowd-funding of the company. Tokens
 23 can be bought directly on the web platform of the company, at different stages of the ICO commonly referred as pre-sale and sale.
 24 Later, the amount of bought tokens can be sold or used in the future to obtain products or services. The portal Tokendata.io has
 25 estimated that until 2017 ICOs raised as much as \$5.3 billion around the world; if we consider venture capitalist, in 2016, they
 26 invested \$71.8 billion in the United States and \$4.3 billion in Europe (National Venture Capital Association and Invest Europe). Start-
 27 ups and existing businesses are turning to alternative sources of capital as opposed to classical channels like banks or venture
 28 capitalists. They can offer the inner value of their business by selling 'tokens', i.e. units of the chosen cryptocurrency, like a regular
 29 firm would do by means of an Initial Public Offering (IPO). When we say cryptocurrency, we refer to a digital currency, a new means
 30 of exchange, the most popular examples of which are Bitcoin and Ethereum. Blockchain (chain of blocks) is the technology at the

basis of a cryptocurrency; it is a Distributed Ledger Technology defined as a distributed, shared, encrypted database that serves as an irreversible and incorruptible repository of information (Wright, De Filippi, 2015). Bitcoin is currently the largest blockchain network followed by, Ethereum, XRP, Litecoin, EOS and Bitcoin Cash (Coinmarketcap.com, March 21, 2019). ICOs favor open-source project development and decentralized business, generating a built-in customer base and positive network effects. They also create a secondary market where tokens can be employed as rewards for using the app of the company or the offered services. This work aims at addressing the specific characteristics of ICOs using relevant variables that play a key role in determining the success of the ICO.

As it stands there is no database with the information we are looking for, thus we have been building and constantly maintaining a dataset that is currently composed of 196 ICOs that occurred between October 2017 and November 2018. The database comprises companies from European countries namely France, Germany, Switzerland, Estonia, Latvia and non European countries such as Russia, United Kingdom, United States, Japan, Singapore, and Australia. The most common sectors in which ICOs operate are: high tech services, financial services, smart contract, gambling platforms, marketplaces and exchanges. The section 4 explains the data collecting workflow.

3 INITIAL COIN OFFERINGS

Most of the ICOs projects are related to the development of a blockchain, the issuance of new cryptocurrencies or somehow related fintech services. ICOs tokens grant contributors the right to access platform services in 68.0% of the cases, governance powers in 24.9% of the cases and profit rights in 26.1% of the cases. The secondary market for ICOs tokens is quite liquid on the first day of trading, and the initial return is large (mean value +919.9% compared to the offer price, median value +24.7%). The success of such decentralized technology lays on the fact that it works without the commitment and the control of a central authority: the blockchain is a Peer-to-Peer technology. A Peer-to-Peer (P2P) system represents a way of structuring distributed applications such that the individual nodes can act as both a client and a server. A key concept for P2P systems is to allow any two peers to communicate with each other in such a way that either ought to be able to initiate the contact (Peer-to-Peer Research Group, 2013). Then, the more a P2P network is distributed, scalable, autonomous and secure, the more is valuable.

All of these precious features have enabled the fast growth of cryptocurrencies not just per se but also as a tool for crow-funding purposes, giving birth to the so-called Initial Coin Offerings. Moreover, what is further fuelling the development of ICOs, according to BIS annual economic report (2018) is the absence of regulation (even if many countries are currently working on it) and, at the moment, there are just a few examples of banning acts (namely China, India, South Korea). Investors buy ICO tokens in the hope of very high returns, sometimes even before the business is put in place, since the corresponding cryptocurrencies (typically Ethereum) can be immediately traded. In the first 6 months of 2018, there have been 440 ICOs, with a peak in May (125) raising more than 10 billion US, where Telegram ICO (Pre-sale 1 2) is by far the most rewarded one with 1.7 billion US (Coinschedule.com, 18 June, 2018). In 2017, the total amount raised by 210 ICOs was about 4 billion US and overcame venture capital funneled toward high tech initiatives in the same period. The first token sale was held by Mastercoin in July 2013 but one of the most successful and still operative is Ethereum which raised 3,700 BTC in its first 12 hours in 2014, equal to approximately 2.3 million dollars at that time.

Recently, there has been a growing literature studying the ICOs drivers aiming to predict their future outcome. A previous study offers an exploratory empiric classification of ICOs and the dynamics of voluntary disclosures. It examines to what extent the availability and quality of the information disclosed can explain the characteristics of success and failure among ICOs and the corresponding projects (Blaseg, 2018). Another important research focuses on the effectiveness of signalling ventures and ICOs projects technological capabilities to attract higher amounts of funding (Fisch, 2018). Momtaz aims at identifying the likelihood and possible timeframe of value creation for investors by combining several factors (financial return, amount of capital raised, listing and delisting alternatives, industry events study etc.) to analyse the ICOs success drivers (Momtaz, 2018c).

Other streams of research concentrate on the impact of managers quality on the ICOs. Momtaz studies the impact of CEOs loyalty disposition and the magnitude of asymmetry of information between managers and investors on ICOs performance (Momtaz, 2018b). Moreover, to remain in the management area, an interesting spark comes from a research specifically directed on CEOs role and effects on ICOs results (Momtaz, 2018a). Finally, another area of studies focuses on the driving factors impacting the liquidity and trading volume of crypto tokens listed after the ICOs. Among those factors have been identified the quality level of disclosed documentation (source code public on Github, white paper published, an intended budget published for use of proceeds), the

community engagement (measured by the number of Telegram group members), the level of preparation of the management (using as proxy the entrepreneurial professional background of the lead founder or CEO), and other outcomes of interest (i.e., the amount raised in the ICO, outright failure - delisting or disappearance, abnormal returns, and volatility) (Howell et al., 2018).

Despite the interest that has been peaked by ICOs and the constantly growing trends, it is worth mentioning that almost half of ICOs sold in 2017 failed by February 2018 (Hankin, 2018). In fact, what should drive more attention to ICOs is the consistent presence of scam activities only devoted to raising money in a fraudulent way. According to Cointelegraph, the Ethereum network (the prevalent blockchain platform for ICOs) has experienced considerable phishing, Ponzi schemes, and other scam events, accounting for about 10% of ICOs (Ethereumscamdb.info, 2017). On the other hand, it is interesting to assess what factors affect the probability of success of an ICO. Adhami et al. in 2018, based on the analysis of 253 ICOs, showed that the following characteristics contribute: the availability of the code source, the organization of a token presale and the possibility for contributors to access to a specific service (or to share profits).

The boom of the ICOs projects and their interesting characteristic brought an important rise of interest from the general audience, many scientific studies have been conducted and published in the last years. Besides the aforementioned Adhami et al. (2018), we should mention the working paper by Zetzsche et al. (2017), that is focused on legal and financial risk aspects of ICOs, moreover a taxonomy is provided and some additional data on ICOs that the authors claim to be continuously updated. Recently, Subramanian in 2018 quoted the ICOs as an example of the decentralized blockchain-based electronic marketplace. The main source of information about blockchain, tokens and ICOs is obviously the Web. Here we can find sites enabling to explore the various blockchains associated to the main cryptocurrencies, including Ethereum's one. We can also find websites giving extensive financial information on prices of all the main cryptocurrencies and tokens, sites specialized in listing the existing ICOs and giving information about them. Often, these sites evaluate the soundness and likeliness of success of the listed ICOs. One of the most popular among these sites is icobench.com, which evaluates all the listed ICOs and provides an API (Application Programming Interface) to automatically gather information on them. ICOs are usually characterized by the following features: a business idea, most of the time explained in a white paper, a proposed team, a target sum to be collected, a given number of tokens to be given to subscribers according to a predetermined exchange rate with one or more existing cryptocurrencies. Nowadays, a high percentage of ICOs are managed through Smart Contracts running on Ethereum blockchain, and in particular through ERC-20 Token Standard Contract (Fenu et al., 2018).

On top of all the characteristics explained so far, there is a further and not yet explored point of interest: the Telegram chats. Telegram is a cloud-based instant messaging and voice over IP service developed by Telegram Messenger founded by the Russian entrepreneur Pavel Durov. In March 2018, Telegram stated that it has 200 million monthly active users - "This is an insane number by any standards. If Telegram were a country, it would have been the sixth largest country in the world (Telegram, 2018)". Telegram is completely free and has no ads, users can send any kind of media or documents and can program messages to self-destruct after a certain period of time. Some characteristics are imposing Telegram among the first social networks, indeed it intentionally does not collect data about where its clients live and what they use the platform for. This is one of the main reason why, according to AppAnnie rankings, Telegram is particularly popular in countries like Uzbekistan, Ukraine, and Russia, where Internet access may be limited or closely monitored by the government. As of October 2017, Telegram was by far the most popular official discussion platform for current and upcoming ICOs, with 75%+ of these projects employing it. This means that retrieving Telegram discussions associated with each and every ICO would produce a huge amount of textual information potentially useful for understanding the chance of success and more interestingly possible signs of fraudulent activities.

4 METHODOLOGY

In this paper we leverage two kinds of information: structured and unstructured ones. Regarding the former, we take advantage of classical statistical classification models to distinguish the status of an ICO that made up of 3 classes, intended as follows:

- Success: the ICO collects the predefined hard cap within the time horizon of the campaign;
- Failure: the ICO does not collect the predefined hard cap within the time horizon of the campaign;
- Scam: the ICO is discovered to be a fraudulent activity with malicious intent during the campaign and described as such by all the platforms we use for data gathering (namely ICObench and Telegram). A robustness check for the scam labeling come by

checking if regulatory bodies announced legal actions against the issuers (e.g. official SEC announcements of legal infringement)

Logistic regression aims at classifying the dependent variable into two groups, characterized by a different status [1=scam vs 0=success or 1=success vs 0=failure] according to the following model:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \alpha + \sum_j \beta_j x_{ij}, \quad (1)$$

where p_i is the probability of the event of interest, for ICO i , $x_i = (x_{i1}, \dots, x_{ij}, \dots, x_{iJ})$ is a vector of ICOs-specific explanatory variables, the intercept parameter α , as well as the regression coefficients β_j , for $j = 1, \dots, J$, are to be estimated from the available data. It follows that the probability of success (or scam) can be obtained as:

$$p_i = \frac{1}{1 + \exp(-(\alpha + \sum_j \beta_j x_{ij}))}, \quad (2)$$

Since the target variable is naturally categorized according to 3 classes, success, failure and scam we extend the aforementioned binary logistic regression to a multinomial one. Such model assesses all the categories of interest at the same time as follows:

$$\ln\left(\frac{p_k}{1-p_K}\right) = \alpha_k + \sum_j \beta_k x_{ij}, \quad (3)$$

where p_k is the probability of k th class for $k = 1, \dots, K$ given the constraint that $\sum_K p_k = 1$.

Considering the textual analysis of Telegram chats, we take advantage of quantitative analysis of human languages to discover common features of written text. In particular the analysis of relatively short text messages like those appearing on micro-blogging platform presents a number of challenges. Some of these are, the informal conversation (e.g. slang words, repeated letters, emoticons) and the level of implied knowledge necessary to understand the topics of discussion. Moreover, it is important to consider the high level of noise contained in the chats, witnessed by the fact that only a fraction of them with respect to the total number available is employed in our sentiment analysis.

We have applied a Bag of Word (BoW) approach, according to which a text is represented as an unordered collection of words, considering only their counts in each comment of the chat. The word and document vectorization has been carried out by collecting all the word frequencies in a Term Document Matrix (TDM). Afterwards, such matrix has been weighted by employing the popular TF-IDF (Term Frequency Inverse Document Frequency) algorithm. Classical text cleaning procedures have been put in place like stop-words, punctuation, unnecessary symbols and space removal, specific topic words addition. For descriptive purposes we have used word-clouds for each and every Telegram chat according to the general content and to specific subcategories like sentiments and expressed moods. The most critical part of the analysis relies on the sentiment classification. In general, two different approaches can be used:

- Score dictionary based: the sentiment score is based on the number of matches between predefined list of positive and negative words and terms contained in each text source (a tweet, a sentence, a whole paragraph);
- Score classifier based: a proper statistical classifier is trained on a large enough dataset of pre-labelled examples and then used to predict the sentiment class of a new example.

However, the second option is rarely feasible because in order to fit a good classifier, a huge amount of pre-classified examples is needed and this represents a particularly complicated task when dealing with short and extremely non conventional text like micro-blogging chats. Insofar, we decided to focus on a dictionary based approach, adapting appropriate lists of positive and negative words relevant to ICOs topics in English language. We employ 3 vocabularies from the R package 'tidytext':

- AFINN from Finn Årup Nielsen;
- BING from Bing Liu and collaborators;

- NRC from Saif Mohammad and Peter Turney.

These lexicons are based on unigrams, i.e., single words, they contain many English words and the words are labeled with scores for positive/negative sentiment and also possibly emotions like joy, anger, sadness, and so forth. The NRC lexicon categorizes words in a binary fashion (“yes”/“no”) into categories of positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. The BING lexicon categorizes words into a binary manner into positive and negative categories. The AFINN lexicon assigns words with a score that runs between -5 and 5 , with negative scores indicating negative sentiment and positive scores indicating positive sentiment. By applying the above described lexicons, we produce for each and every ICO a sentiment score as well as counts for positive and negative words. All these indexes are used as additional predictors within the logistic models.

5 DATA

In this paper, we examine 196 ICOs starting from January 2017 till November 2018. For each project, we gather information from web-based sources, mainly rating platforms such as icobench.com, TokenData.io, ICO Drops.com, CoinDesk.com and project’s websites. The process of building up the ICOs data set reflects the main phases that an ICO follows to be launched: from the birth of the business idea, the team building, the purpose of the tokens, the technical requirements (white paper), the promotion and the execution phase.

5.1 Collection of structured data

The first step in collecting data about each project is to gather information from the most used ICO related platforms as Icobench, TokenData, Coinschedule or similar. During such phase, we look for general characteristics such as the name, the token symbol, start and end dates of the crowdfunding, 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 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 variables set, continuous and categorical data, show us that the main area of origin of the projects is Europe with the highest percentage in Switzerland and Germany. The Switzerland peak is due to the national regulator approach - FINMA (the Swiss Financial Market Supervisory Authority) - which on 16 February 2018 issued clear guidance on the status of ICOs. FINMA does not categorise payment or utility tokens (provided they are not used for investment) as securities. All other tokens are categorised as securities and are subject to securities regulation. To legally issue an equity/asset token, authorization from FINMA should be sought, and appropriate compliance measures (know your customers (KYC) and anti-money laundering (AML)) must be taken. If a debt token can be classified as a deposit, then unless specific exceptions apply, a banking license is needed prior to the ICO. In the fragmented regulatory framework, this is one of the so-called “crypto-friendly” countries, that attract worldwide investors.

The presence of a team of experts as a figure of “advisors” that follows the stages of development are helpful in qualifying the ICO as more reliable. On the development of the dataset the research focused also on assessing the number of advisors for each ICO, checking their educational background and marking as a variable of interest the presence of a Ph.D. that attests a high degree of education.

The evolution of the classic Business Plan that we observe when we analyse the idea of a start-up, is called White Paper. The business plan is the document that illustrates the strategic intentions and the management of competitive strategies of the company, the evolution of key value drivers and the economic and financial results. The drawing up of the operational plan has the aim of achieving different subjects involved in the business. The content of the business plan should not be overlooked, it must be the most possible schematic and of intuitive interpretation. The feature that distinguishes a good plan is the clarity, the synthesis and the professional description of the project workflow. The WP (white paper) therefore fulfills these functions and, in our analysis, played a vital role in the statistical analysis in terms of presence or absence of it. The graphics quality with which it is produced is also important, the data contained within it and the description of the team’s components.

5.2 Collection of Unstructured data

Social channels are more personal than every database, rating platform or websites, so they are a way to reach a wide range of users, to update them constantly about the evolution of the project and in the end to create a trusty environment that can finalize in a successful crowdfunding activity. In order to conduct the textual analysis, we enrich our database with the social channels data, such as the presence of a channel, the numbers of users as a proxy of the community engagement and as mentioned in the introduction the textual chat, retrieved in reverse until the creation of the chat. The most used social channels are Telegram, Twitter, Facebook, Bitcointlak, Medium, while Linkedin, Reddit and Slack are not frequently used.

In crowdfunding projects the entrepreneur and the community in which is embedded works as a strong control for the attractiveness of a business. Some studies have investigated the social network community and the entrepreneurial activity finding out that the amount of capital collected in crowdfunding is heavily dependent on the range of social networks the entrepreneurs belong to (E. Mollick, 2014).

With regards to the entrepreneurial dimension, we investigate the team components, pointing out that the members checked until now are almost 1000, with a median size of 7 for project. For each team member we checked general information related to the social engagement, looking for the Linkedin channel activity (48 % of them do not have an individual page), the numbers of connections, the job position in the project and the academic background. Moreover the presence of advisors can play a crucial role in ensuring the reliability of an ICO, provided a wise choice of such advisors. The same applies to institutional investors doing due diligence on a potential venture. In collecting our data, we focused on the academic background and the current area of expertise of the declared advisors.

As it concerns the unstructured data, insightful information can be derived by the white papers in terms of quality of the technical report and specific content. A white paper is a summary report that provides detailed information about the project, its originality and the benefits it can give to investors and users, about the technological features, team behind the project, project's background and future plans. Besides all the above information, we collect Telegram chats associated to each ICO (if available) and apply all the text analytic techniques to produce a sentiment based score.

In Table 1 we report the complete list of collected and employed variables.

6 EMPIRICAL EVIDENCE

In this section we report our main results obtained from classification analysis and textual analysis. In this regard, in table 2 and table 4 we report results respectively for logistic regression on Success/Failure (class 2 variable) and for multilogit regression estimated on failure (f) and scam (sc) compared to success as baseline. Regarding the first model, in table 2 we report the final configuration after several stepwise selection steps¹. The reader can see that the only two relevant dummy variables are: the presence of a white paper (Paper_du) and of a Twitter account (tw). Both present positive coefficients showing their impact on increasing the probability of success of an ICO. It should be stressed that the influence of Twitter channel is much higher than the presence of a white paper, indeed if we calculate the associated odds ratio we would get respectively 11.94 and 3.85. In other words, if the ICO has a Twitter account the probability of success is almost 12 times higher (almost 4 times higher for the white paper). Regarding the three continuous variables, number of elements of the team (Nr_team), number of advisors (Nr_adv) and scaled sentiment score based on NRC lexicon (Sent_NRC_sc), they are all highly significant and again positive suggesting that increasing people and advisors in the team has a positive impact. Regarding the sentiment, we notice a particularly high positive value, stressing the importance of the perception of possible investors which interact with the ICO proposer by means of a social media, namely Telegram.

To further evaluate such configuration, we have explored the VIF index that accounts for the level of multicollinearity brought by each and every variable. The VIF results for the two model configurations are reported in table 3 (logistic) and 5 (multinomial), with useful insights in defining the lack of multicollinearity². Therefore in table 3 we can see low values for the VIF index associated to the estimated logistic model (given in table 2). The reader can easily notice that there is not any multicollinearity effect, making robust the model. Moreover, reported performance indexes, namely AIC and pseudo R^2 , present good values above 50%.

In table 4, we report results for fraudulent and scam ICOs compared to successful ones, on the basis of a multilogit regression. Looking at the estimated parameters, we can infer that the patterns are different. The presence of a website has a positive impact on the probability of being a successful ICO and not a scam. In other words, the absence of this characteristic is a driver of

¹ The full model is available in table 6 in the appendix and it evidently contains several not significant variables.

² In table 7, we have VIF index obtained from the full model and there are high values for some variables, specifically those related to sentiment analysis.

Table 1. Explanatory variables

class0	f=failed, sc=scam su=success
class1	0=scam, 1=failed+success
class2	0=failed, 1= success
w_site	Website (dummy)
tm	Telegram (dummy)
w_paper	White paper (dummy)
usd	presale price in USD
tw	Twitter (dummy)
fb	Facebook (dummy)
ln	Linkedin (dummy)
yt	Youtube (dummy)
gith	Github (dummy)
slack	Slack (dummy)
reddit	Reddit (dummy)
btalk	Bitcointalk (dummy)
mm	Medium (dummy)
nr_team	Number of Team members (quantitative)
adv	Existence of advisors (dummy)
nr_adv	Number of advisors (quantitative)
project	Official name of the ICO (categorical)
nr_tm	Number of users in Telegram (quantitative)
tot.token	Number of Total Tokens (quantitative)
Pos_Bing	Standardized nr. of positive words for BL list (quantitative)
Neg_Bing	Standardized nr. of negative words for BL list (quantitative)
Sent_Bing	Standardized sentiment for BL list (quantitative)
Pos_NRC	Standardized nr. of positive words for NRC list (quantitative)
Neg_NRC	Standardized nr. of negative words for NRC list (quantitative)
Sent_NRC	Standardized sentiment for NRC list (quantitative)

scam activity suspects. Instead the website does not differentiate successful from failures ones. With regards to the presence of advisors and of a white paper, both the variables are significant in differentiating fraudulent from successful ICO, confirming results of logistic regression. No statistical significance for fraudulent ICOs. Lastly, variable on the sentiment score is relevant and with negative sign for both the classes, in other words an increasing in the sentiment causes an increasing in the probability of success when we consider both failed and fraudulent ICOs.

In this regard, we should stress that the incidence of scam ICOs in our database is extremely low, this due to the fact that collecting information about such ICOs is particularly complex. Most of the information is completely deleted from the Web as soon as the activity is recognized as illegal and/or fraudulent. The overall model performance, assessed again in terms of AIC and pseudo R^2 , is pretty good although inferior to the previous one.

In table 5 we also report VIF index, so to check the absence of multicollinearity in the reported model. Please note that, multilogit model reported in table 4 is a final configuration obtained through stepwise selection. The full models are available in the Appendix section ³.

7 DISCUSSION AND CONCLUSIONS

Initial coin offerings (ICOs) are one of the several by-products of the cryptocurrencies world. IPOStart-ups and existing businesses are turning to alternative sources of capital as opposed to classical channels like banks or venture capitalists. They can offer the inner value of their business by selling 'tokens', i.e. units of the chosen cryptocurrency, like a regular firm would do by means of an . The investors, of course, hope for an increase in the value of the token in the short term, provided a solid and valid business idea typically described by the ICO issuers in a white paper. However, fraudulent activities perpetrated by unscrupulous actors are frequent and it would be crucial to highlight in advance clear signs of illegal money raising.

³ Full multilogit regression model is available in table 8 in the appendix and in table 9 the associated VIF index table.

Table 2. Logistic Regression Results on Success/Failure ICOs

<i>Dependent variable:</i>	
	class2
tw	2.481* (1.381)
Paper_du	1.351** (0.635)
nr_adv	0.461*** (0.135)
nr_team	0.233*** (0.088)
Sent_NRC_sc	2.187*** (0.595)
Constant	-3.601** (1.458)
Observations	196
Akaike Inf. Crit.	89.41
McFadden pseudo R ²	0.63
McFadden Adj. pseudo R ²	0.57
Cox & Snell pseudo R ²	0.49
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

Table 3. VIF index for logistic regression model

tw	Paper_du	nr_adv	nr_team	Sent_NRC_sc
1.229	1.033	1.067	1.053	1.228

In this perspective, ICOs analysis can be considered a very particular type of fraud detection activity. However, in our opinion fraud detection presents some specificity that prevent us from entailing ICOs related problems as a proper instance of fraud detection. In particular, our data are not flowing in such huge amount from an on-line system as typically happens with credit card payments or banks transactions. Typical fraud detection approaches, as in Maheshwara et al. (2018), aim at discovering, almost in real times, fraudulent financial activities based on transactional data that ideally should be blocked as soon as possible. ICOs instead are characterized by a slow process of engagement of the prospect clients and establishment of consensus that goes through Telegram chats (if available), white paper and website. That being the case, we would suggest to label this specific stream of research as FinTech Fraud detection with all the relative specificity.

While analyzing success vs failure dynamic with a classification model is relatively easy since the incidence of the two classes is almost equal (50-50), it is much more complicated to highlight the key aspects that could witness a fraudulent activity since, in the last 3 years, only few scam events have been reported. In our sample made of 196 ICOs (data collection still active) we have 10 scam ICOs and we fit a multilogit regression model for comparing scam and failed ICOs towards successful ones. Results tell us that the presence of a website has a positive impact on the probability of not being a scam but does not have any impact on failed ones. In terms of sentiment expressed on Telegram chats, the impact appears to be negative both on the scam and failed ICOs. This suggests that monitoring in real time Telegram chats could represent a valid mean for collecting signs of possible problems within the ICOs. If instead, we compare Successful ICOs against Failed ones, we find that the presence of a White Paper and of a Twitter account show positive coefficients.

Table 4. Results from multilogit regression: failure and scam compared to success

	<i>Dependent variable:</i>	
	sc (1)	f (2)
Oweb_dum	−1.962** (0.977)	0.093 (0.773)
adv_dum	−0.899 (0.809)	−1.707*** (0.571)
Paper_du	−0.728 (0.915)	−2.158*** (0.657)
Sent_NRC_sc	−1.390* (0.731)	−2.606*** (0.703)
Constant	−0.628 (0.997)	−0.572 (0.925)
Akaike Inf. Crit.	166.339	166.339
<i>Pseudo R square</i>	McFadden 0.43 - McFadden Adj. 0.36- Cox & Snell 0.44	

Note: *p<0.1; **p<0.05; ***p<0.01

Table 5. VIF index for multilogit regression model

Oweb_dum	adv_dum	Paper_du	Sent_NRC_sc
3.656	2.317	3.607	3.870

Regarding the three continuous variables, number of elements of the team, number of advisors and sentiment score based on NRC lexicon, they are all highly significant and positive suggesting that increasing people in the team and advisors has a positive impact. Regarding the sentiment, we notice a particularly high positive value, stressing the importance of the perception of possible investors which interact with the ICO proposer by means of a social media.

The paper will be improved in the future by increasing the size of the sample and exploring alternative approaches for textual analysis with specific attention to sentiment analysis. We aim at producing a more refined and tailored sentiment score for each ICO, improving and increasing the vocabulary of words. Specifically regarding the textual analysis an alternative approach that we could use is the combination of words as in Bolasco and Pavone, (2010).

As a final remark, authors are aware of the limits of the paper mainly due to the size of the sample. However, given the still limited literature in this field with no reference to the power of textual information collectable through Telegram chats, this contribution represents a step ahead in the process of understanding the ICOs phenomenon. Furthermore a different approach would be to study the trends of the ICOs by combining the available information from specialized websites on fraudulent activities (such as cyphertrace.com and deadcoin.com) and rating websites for the active projects.

8 REFERENCES

- [1] Adhami S., Giudici G., Martinazzi S.,(2017). Why Do Businesses Go Crypto? An Empirical Analysis of Initial Coin Offerings, available at SSRN: <https://ssrn.com/abstract=3046209>
- [2] BIS Annual Economic Report,(2018). V. Cryptocurrencies: looking beyond the hype,
- [3] Bolasco S. and Pavone P. (2017). Automatic dictionary-and rule-based systems for extracting information from text. In Data Analysis and Classification (189-198). Springer Berlin Heidelberg.

- [4] Blaseg, D.,(2018). Dynamics of Voluntary Disclosure in the Unregulated , Market for Initial Coin Offerings, available at SSRN: <https://ssrn.com/abstract=3207641> or <http://dx.doi.org/10.2139/ssrn.3207641>
- [5] Calabrese R., Marra G., Osmetti S.A. (2016). Bankruptcy Prediction of Small and Medium Enterprises Using a Flexible Binary Generalized Extreme Value Model. *Journal of the Operational Research Society*, 67(4), 604-615
- [6] Coinschedule, (2018). "Cryptocurrency ICO Stats 2018" available at: <https://www.coinschedule.com/stats.html> [Accessed June 18, 2018]
- [7] Coinmarketcap, (2018). "All coins" available at: <https://coinmarketcap.com/coins/views/all/> [Accessed June 18, 2018]
- [8] Ethereumscamdb, "Ethereum Scam Database" available at: <https://etherscamdb.info/> [Accessed June 18, 2018]
- [9] Falk, M., Haler, J., and Reiss, R. (2010). *Laws of Small Numbers: Extremes and Rare Events*. Springer.
- [10] Fenu, G., Marchesi, L., Marchesi, M. and Tonelli, R., (2018). The ICO phenomenon and its relationships with ethereum smart contract environment. In 2018 International Workshop on Blockchain Oriented Software Engineering (IWBOSE) (pp. 26-32). IEEE.
- [11] Fisch C.,(2019). Initial Coin Offerings (ICOs) to Finance New Ventures. *Journal of Business Venturing*, 34(1), 1–22.. Available at SSRN: <https://ssrn.com/abstract=3147521> or <http://dx.doi.org/10.2139/ssrn.3147521>
- [12] Hankin A.,(2018).Nearly half of all 2017 ICOs have failed. *Fortune*, [Retrieved May 4, 2018].
- [13] Howell S. T., Niessner M., and Yermack D. (2018). Initial coin offerings: Financing growth with cryptocurrency token sales (No. w24774). National Bureau of Economic Research.
- [14] Kotz, S. and Nadarajah, S. (2000). *Extreme Value Distributions. Theory and Applications*. Imperial College Press, London.
- [15] Peer-to-Peer Research Group, (2013). "Peer-to-Peer Research Group" available at: <https://irtf.org/concluded/p2prg> [Accessed 18 June 2018]
- [16] Maheshwara Reddy C., M.Saiteja, B.Shashank, Mrs.Dhikhi (2019). Financial Fraud Detection Using Machine Learning. *IJIRT*, Volume 6 Issue 6, ISSN: 2349-6002
- [17] Mollick E.,(2014). The dynamics of crowdfunding: an exploratory study, *J. Bus. Venturing*, 29 (1) , pp. 1-16
- [18] Momtaz P. P., (2018).CEO Emotions and Underpricing in Initial Coin Offerings. Working paper, UCLA Anderson
- [19] Momtaz P. P., (2018). Initial Coin Offerings, Asymmetric Information, and Loyal CEOs. SSRN. <https://doi.org/http://dx.doi.org/10.2139/ssrn.316709>
- [20] Momtaz P. P. (2018). Initial Coin Offerings. SSRN. <https://ssrn.com/abstract=3166709>
- [21] Subramanian H. (2018). Decentralized Blockchain-Based Electronic Marketplaces, *Comm. ACM*, vol. 61, pp. 78-84.
- [22] Zetsche D.A. , Buckley R.P., Arner D.W., Föhr L.,(2017). The ICO Gold Rush:It's a Scam, It's a Bubble, It's a Super Challenge for Regulators, Univ.Luxembourg Law Working Paper No. 11/2017, available at SSRN: <https://ssrn.com/abstract=3072298>, November 2017.
- [23] Wright A. , De Filippi P.,(2015). Decentralized Blockchain Technology and the Rise of Lex Cryptographia, available at SSRN: <https://ssrn.com/abstract=2580664> or <http://dx.doi.org/10.2139/ssrn.2580664>

9 APPENDIX

Table 6 Logistic regression full model

Variables	Dependent variable:
	Coeff. & s.e.
	class2
Oweb_dum	0.096 (1.155)
Code_dum	−0.416 (0.923)

Continuation of Table 6	
Variables	Coeff. & s.e.
Telegram_du	−0.492 (1.073)
tw	2.858 (1.773)
fb	0.107 (0.817)
ln	−1.236 (0.927)
yt	1.495 (1.002)
gith	−0.432 (0.865)
slack	−0.915 (1.105)
reddit	−0.067 (0.902)
btalk	−0.611 (0.811)
ew	10.598 (1,648.456)
mm	0.997 (0.902)
Neg_Bing_sc	−5.293 (3.846)
Pos_Bing_sc	2.730 (1.697)
Neg_NRC_sc	4.454 (4.058)
Pos_NRC_sc	0.749 (1.829)
Paper_du	1.487 (1.028)
nr_adv	0.582*** (0.197)
nr_team	0.268** (0.124)
Constant	−3.649* (2.018)
Observations	196
Log Likelihood	−31.727
Akaike Inf. Crit.	105.454
Note: *p<0.1; **p<0.05; ***p<0.01	

Table 7. VIF index for the full model

Oweb_dum	Code_dum	Telegram_du	tw	fb
2.034	1.394	2.499	1.572	1.419
ln	yt	gith	slack	reddit
2.058	1.649	1.805	1.348	2.009
btalk	ew	mm	Neg_Bing_sc	Pos_Bing_sc
1.585	1.000	2.003	73.564	9.982
Neg_NRC_sc	Pos_NRC_sc	Paper_du	nr_adv	nr_team
70.371	12.830	2.255	1.703	1.809

Table 8 Multilogit regression full model

Variables	Dependent variable:	
	Coeff. & s.e.	
	sc (1)	f (2)
Oweb_dum	-4.568** (2.261)	1.602 (1.613)
Code_dum	-1.051 (1.745)	1.312 (1.222)
Telegram_du	3.146 (2.809)	0.275 (1.340)
tw	-8.198** (4.160)	-5.940 (3.974)
fb	3.333 (2.577)	-0.273 (0.933)
ln	1.631 (1.945)	1.448 (1.185)
yt	-2.427 (2.494)	-1.007 (1.241)
gith	0.693 (1.713)	-0.044 (1.057)
slack	-0.834 (2.469)	1.616 (1.351)
reddit	2.058 (2.338)	-0.551 (1.133)
btalk	0.641 (1.713)	0.774 (0.927)
ew	-0.849 (137.854)	-2.641 (34.907)
mm	-1.703 (2.317)	-0.229 (1.214)
Neg_Bing_sc	15.599 (13.027)	-6.986 (8.735)
Pos_Bing_sc	-4.379 (2.882)	-10.411 (12.921)
Neg_NRC_sc	-14.856 (15.289)	5.266 (5.423)

Continuation of Table 8		
Variables	Coeff. & s.e.	
	sc (1)	f (2)
Pos.NRC_sc	−1.593 (3.984)	6.095 (8.983)
Paper_du	−1.457 (1.744)	−2.143* (1.243)
nr_adv	−0.564 (0.377)	−0.715*** (0.254)
nr_team	−0.047 (0.106)	−0.454*** (0.148)
Constant	0.369 (4.177)	4.372 (3.062)
Akaike Inf. Crit.	163.240	163.240
Pseudo R square	McFadden 0.46 - McFadden Adj. 0.37 - Cox & Snell 0.46	
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 9. VIF index multionomial full model

Oweb_dum	Code_dum	Telegram_du	tw	fb
15.583	3.165	9.422	126.167	5.164
ln	yt	gith	slack	reddit
3.883	2.276	4.194	2.408	4.640
btalk	ew	mm	Neg_Bing_sc	Pos_Bing_sc
3.443	1.007	6.435	575.279	1,293.780
Neg_NRC_sc	Pos_NRC_sc	Paper_du	nr_adv	nr_team
241.850	770.032	7.955	3.908	8.418

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

The paper is the product of full collaboration among the authors, however PC has inspired the idea, the methodology and wrote section 2, 4 and 7, AMT run the analysis and wrote section 1, 3, 5 and 6.

FUNDING & ACKNOWLEDGMENTS

This research has received funding from the European Union's Horizon 2020 research and innovation program "FIN-TECH: A Financial supervision and Technology compliance training program" under the grant agreement No 825215 (Topic: ICT-35-2018, Type of action: CSA) The current version of the work is substantially based on refining a working paper(not reviewed) published by the same authors in DEM Working Paper Series "ICOs success drivers: a textual and statistical analysis". We acknowledge Federico Campasso that technically helped us in collecting the data and Prof. Silvio Vismara for the useful advises on the work.