

JULY 10, 2018



MSc in Management, Digital Business Major Thesis

**ANALYSIS OF SUCCESS FACTORS
FOR INITIAL COIN OFFERINGS AND
AUTOMATISATION OF WHITEPAPER
ANALYSIS USING TEXT-MINING
ALGORITHMS**

BUGRA CETINGOK

BERNAT SERRA DEOLA

Tutor: LAURENT BENICHO

Table of contents

A. Abstract	2
B. Introduction to Blockchain	3
1. Use cases of Blockchain	4
C. Understanding ICOs.....	6
1. Context.....	6
2. What is an ICO.....	6
3. History and evolution of ICOs.....	7
4. Creation of an ICO	12
D. Investing in ICOs.....	15
1. Main differences between an ICO and an IPO.....	15
2. Advantages of investing in ICOs.....	17
3. Risks of investing in ICOs	18
4. How to evaluate an ICO	20
5. The Definition of Success for an ICO	22
E. Predicting the Financial Success of ICOs Based On Data and the Whitepaper Analysis	23
1. Context.....	23
2. Data collection.....	24
3. Data analysis.....	24
4. Text mining algorithm	27
F. Conclusion.....	39

A. ABSTRACT

Executives and visionaries have already explored many use cases of blockchain. However, the discovery of using cryptocurrencies, which are created by using blockchain, as means of crowdfunding in the form of initial coin offering (ICO) had a huge impact on shaping the way people invest in 2017. ICOs became one of the first use cases of blockchain that changed a highly regulated finance industry and they have been in high demand since the success of Ethereum project.

Investing in ICOs allows small individual investors interested in tech to indirectly participate and get involved in revolutionary projects that aim to disrupt the state of the art of the current industries and set new social and economic standards.

Linked to the mid and long-term success of the blockchain projects and its implementations into real practical applications, crypto-tokens can be used to purchase the services offered and in the case of high popularity and market trust, they can be also used for third-party purchases.

The counterparts of investing in ICOs are strongly linked with its benefits. The myriad of opportunities, the inexistent barriers of entry and the early-stage of the involved projects make them extremely attractive for small investors. However, these investments offer little or no financial guarantees and are subject to important uncertainties about the feasibility of the business model and the background of the developing team. Many impostors blossomed to benefit from the enormous demand, especially over the course of 2017.

In this paper, we put forward the points to be considered before investing in ICOs and supported our assumptions by conducting data analysis on 106 ICOs and surveys on 50 amateur investors. Based on our study, we realized that one of the major pain points is to analyse white papers, as it is both time consuming and amateur investors struggle to understand the maturity, the potential use cases and the value proposition of the projects. In this context, we developed a text-mining algorithm with accuracy between 95%-100% by using R Studio and its natural language processing libraries to apply machine learning techniques on white papers in order to automatize the investment decision based on the scope of the project.

B. INTRODUCTION TO BLOCKCHAIN

“Blockchain refers to a distributed, encrypted database, which is a public depository of information that cannot be reversed and is incorruptible.”¹ The distributed ledger theory builds the constitution of Blockchain technology, where the ledger might indicate any kind of digital information or transaction and is stored and processed by a distributed computer network, in which each computer is called as a ‘miner’.

Blockchain is characterized by the following principles:

- It is decentralized and peer-to-peer
- Certain users can write or read the data, as digital signatures are employed, and the users have to have private keys, which are cryptographically created, to access the data (blocks) they own
- Its architecture makes it almost impossible to change historical records and provides easy detection of fraud attempt
- ‘Proof of Work’, which is a security protocol and first proposed by Moni Naor and Cynthia Dwork in 1993, is employed to enhance the network security. It allows the decentralized network to arrive to a consensus about the state of transactions²
- Miners are rewarded based on Proof-of-Stake protocol, a method by which a Blockchain network aims to achieve distributed consensus²

Apart from the exclusion of third parties, which drastically reduces the cost of transactions and fastens the verification process, to provide the trust mechanism, the Blockchain technology brings about many other advantages. Blockchain does not rely on a single central server, but a distributed network of computers, which provides more reliability and durability, as the entire network would not fail because of attacks or maintenance issues. Furthermore, it strengthens users in means of transparency and integrity since the users can control their transactions.

¹ Morabito, V.: Business Innovation Through Blockchain

² Antonopoulos, A.: Mastering Bitcoin

On the other hand, the implementation of Blockchain technology in many industries still faces plenty of challenges, such as regulations, privacy, lacking a well-established developer's community, operational risks, initial implementation cost, cultural acceptance and many other. For instance, financial transactions are highly regulated by national governments and new regulations about the adoption of Blockchain by financial institutions should be designated, so that financial institutions could invest in the technology confidently. Cyber security issues are also deterrent for people to share personal information and makes the technology fragile against the regulations.

To sum up, Blockchain is a nascent, promising technology which can have an impact on transforming businesses, increasing the prosperity of the society and shaping the future. Its potential to decentralize the control of powerful institutions and to disrupt the traditional businesses makes the technology face potential challenges such as regulative issues. However, it is very likely that people will be embracing Blockchain, as they gradually discover and understand how it is the use cases of the Blockchain technology can help boost the trust mechanism between parties, increase the efficiency of a business and augment the prosperity of the society.

1. USE CASES OF BLOCKCHAIN³

A recent IBM study over 3000 executives indicates that more than 80% are actively using or planning to implement the blockchain technology into their daily operations. The outcome is not surprising, as the use cases of blockchain are vast and promising to improve the operations of businesses. The main value proposition of Blockchain is to reduce the costs by eliminating the intermediary or middlemen, improve the efficiency and increase the security. In this section, we will analyse how the fundamentals of blockchain will be implemented in different industries and creating value.

Payment industry is one of the first industries which started to explore the opportunities in implementing blockchain in their operations. Cryptocurrencies

³ Blockchain in business: <https://lisk.io/academy/blockchain-business/blockchain-in-business>

are expected to be used within the context of e-commerce, payment or money transfer as an alternative to credit cards. Blockchain payment options will also improve the payment options by offering pay per article, pay per minute or pay per view options in advertising and the online media sector rather than pay per month method, as they lower the costs and reduce the confirmation period of transactions.

Furthermore, commercial giants are constantly looking for ways to increase transparency and price efficiency in their supply chain network. As a result, big tech companies, such as IBM, are significantly investing in blockchain technology to create SaaS solutions, so that their clients get better insights about their supply chain network. Applying blockchain can address to the problems in supply chain by tracking a permanent history of a product in a more secure way of real-life data tracking, eliminating the middlemen.

Using blockchain as a source of trust can positively impact the accounting processes. Especially, the decentralized ledger heralds full automated audits. It helps organizations saving huge amount of labour work by providing an immutable alternative. It also decreases the need for organizations to validate authenticity.

Moreover, integration of blockchain in cloud storage help companies maintain their competitive advantage in their business intelligence since the cryptography can be used to protect the network of interconnected databases and IoT devices from external attacks.

C. UNDERSTANDING ICOS

1. CONTEXT

This chapter gathers the key concepts and definitions from the most popular tech blogs and financial and crypto sites to provide the reader with a global understanding of ICOs and how do they work. Hence, the following part of the study covers from the basic concepts and history of Initial Coin Offerings to financial speculation and best investment practices, including the temporal evolution of the whole process as well as its costs, factors of success and the inherent risks of such an investment.

2. WHAT IS AN ICO

An Initial Coin Offering, better known as an ICO, is a way to crowdfund projects used by blockchain-based startups through the issuance of a token. In the future, when the blockchain platform of the project is created, these tokens will be used to access its resources, products or services. For example, tokens of the project Filecoin gave access to storage space, the ones from OpenMined were exchanged for datasets and, using a more popular example, Ethers are used as a transaction fee charged to its users when creating smart contracts in the Ethereum platform.

Therefore, tokens are created fundamentally by and for developers, who will use them to develop their own blockchain applications on top these platforms thanks to the open-source nature of the projects, in which the platform code is available to the community.

However, despite having such a specific role within the blockchain ecosystem, the value of tokens is totally different when their holders act as speculative investors. That's when these tokens are traded in cryptocurrency exchanges for other tokens that give access to FIAT currencies, such as Bitcoin and Ethereum. In such cases, that represent the vast majority of ICO investments, tokens become cryptocurrencies and, as a matter of fact, mere currencies.

3. HISTORY AND EVOLUTION OF ICOs

a. J.R. Willett, the pioneer

It was in 2013, during the Bitcoin Conference held in San Jose, that the first ever ICO was announced in video by the panellist J.R. Willet, a Seattle-based software engineer. But first, let's move back to January 2012, when Willet published a white paper titled *The Second Bitcoin White Paper* on the Bitcoin Talk forum. In it, he claimed that the existing bitcoin network could be used as a protocol layer on top of which new currency layers with new rules could be built. Then, these new protocol layers would be implemented by newly built software that would richly reward the early adopters of the new protocol.

Back to the San Jose Bitcoin Conference in 2013, Willet described his idea, in video⁴, as follows:

If you wanted to, today, start a new protocol layer on top of Bitcoin, a lot of people don't realize, you could do it without going to a bunch of venture capitalists and instead of saying, hey, I've got this idea, you can — you're familiar with Kickstarter I assume? Most of you? You can actually say, okay, here's my pitch, here's my group of developers — there's a lot of developers in this room. If you get a bunch of trustworthy guys together that people have heard of and say, okay, we're going to do this. We're going to make a new protocol layer. It's going to have new features X, Y and Z on top of bitcoin, and here's who we are and here's our plan, and here's our bitcoin address, and anybody who sends coins to this address owns a piece of our new protocol. Anybody could do that. And I've been telling people this for at least a year now because I want to invest in it. I don't have a ton of coins, but that's where I want to invest my coins. And I've yet to find somebody who wants my coins. Does anybody in this room want my bitcoins because I want to—

I'll take them! – someone shouted.

⁴ https://www.youtube.com/watch?v=4bMf4xZg_4U&feature=youtu.be&t=4m19s

Without knowing it, in 2012 Willet described for the first time the concept that, 5 years later, in 2017, would be built on Ethereum (not on Bitcoin) through the process described also by himself in 2013, the Initial Coin Offering.

Mastercoin, the first ICO

Later that 2013, Willet launched the first ever ICO, Mastercoin, currently known as Omni. He realized that he could create a coin on top of Bitcoin that buyers would automatically own in exchange of Bitcoins to fund its development. After one year of promoting his idea without much success, he did what he called an “initial distribution” by himself, motivated by the idea of proving that it could actually work.

The crowdfunding process took one month and raised 5,000 Bitcoins, \$500,000 at the time, and despite Mastercoin proposed innovative ideas like decentralized commerce and exchange, Willet realized that the success was caused by the excitement of the people towards this new process: “publishing a paper and anonymous people sending you money”.

b. Ethereum

A few months after the launch of Mastercoin other projects started doing their own ICOs, the most famous of which was Ethereum, in July 2014. The ICO process lasted 42 days and within the first 12 hours it had already raised 3,700 BTC (\$2.3 million at the time). With a total funding of 31,500 BTC (\$18.4 million at the time), it is still the second most successful ICO and the 6th highest funded crowdfunding project so far.

Besides its massive financial success, with a return on investment of more than 200,000% for those who participated in the token crowd sale, Ethereum also got ahead of Bitcoin to become, as of February 2018, the leading blockchain platform for ICOs with more than 80% of applications being built on top of its protocol layer⁵.

⁵ "Ethereum Scam Database". ethereumscaMDB.info. Retrieved 2017-08-20

Besides being a developer friendly cornerstone, the success behind Ethereum lies in the fact that democratized ICOs through its ERC20 protocol standard. This protocol layer served as a template to implement all basic functions of a tradeable token and guarantee the tradability on secondary exchanges with other tokens based on the same protocol.

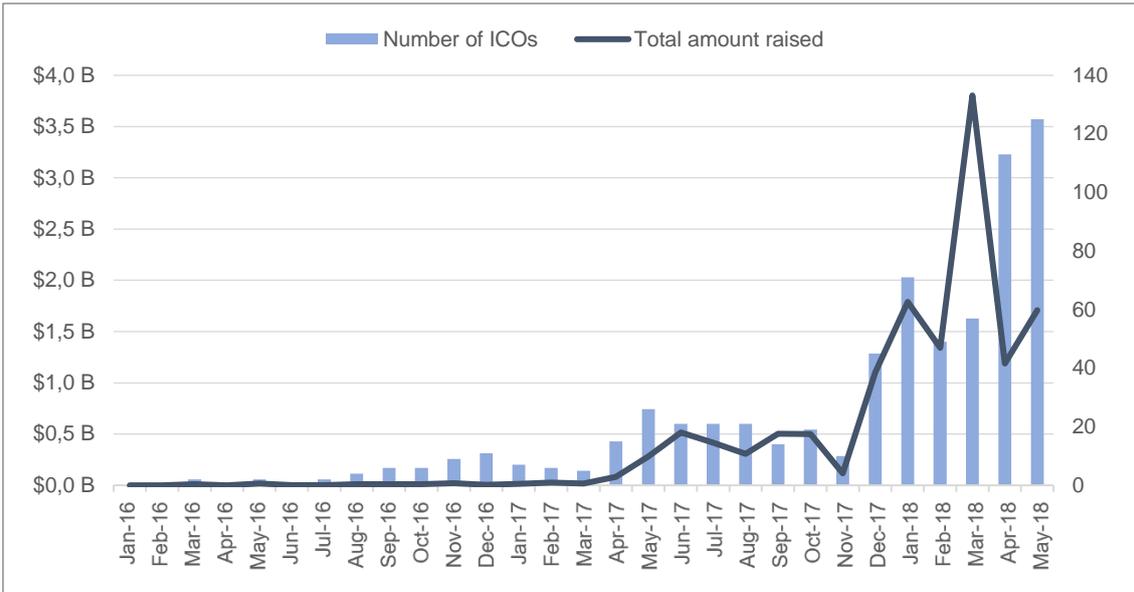
At present, Ethereum has become a decentralized platform that runs smart contracts, applications that run exactly as programmed without any possibility of downtime, censorship, fraud or third-party interference. These apps run on custom built blockchains, shared global infrastructures that can move value around and represent the ownership of property. This enables developers to create markets, store registries of debts or promises, move funds in accordance with instructions given long in the past (like a will or a futures contract) and many other things that have not been invented yet, all without a middleman or counterparty risk.

c. 2017 triggered the current ICO bubble

The ICO market boomed in 2017, with 210 ICOs raising a total of \$3.9 B⁶, a figure 40 times higher than the one raised in 2016, when 43 ICOs were created. Far from being a transient trend, the volume of ICO investment keeps growing exponentially, with \$9,8 B raised so far in 2018 and 415 ICOs created.⁷

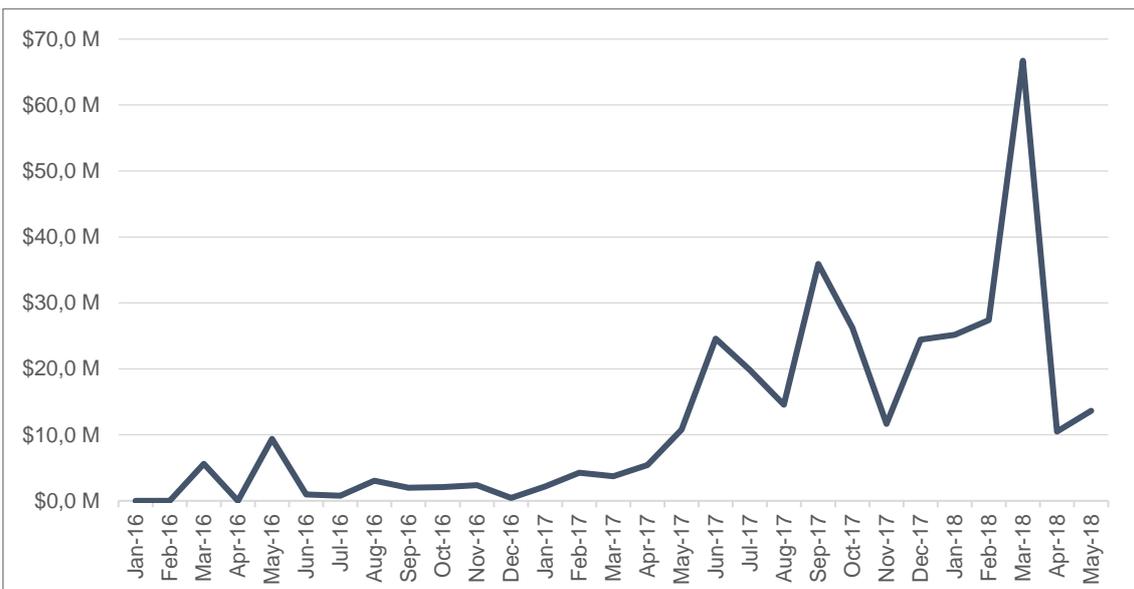
⁶ Valuation based on BTC exchange rate at the time of the ICO closing date. Source: CoinSchedule

⁷ Source: CoinSchedule



Monthly evolution of ICOs and total funds raised from January'16 to May'18. Source: CoinSchedule

There is a massive increase of almost 4,000% in the total proceeds raised from 2016 to 2017, and more than 2,200% between Jan.'17-May'17 and Jan.'18-May'18. This growth is driven by the amount raised per ICO rather than by the volume of ICOs conducted. The following graph illustrates this growing trend, with an average amount raised per ICO of \$67 M during March 2018, followed though by a sharp decrease next month.



Monthly evolution of the average amount raised per ICO. Source: CoinSchedule

Whereas the top ICO of 2016 raised \$16,4 M, the 10th most successful ICO of 2017 raised \$53,0 M, with the first one raising \$258 M. Year-to-date, Telegram is still the top ICO having raised \$1,7 B through 2 pre-sales.

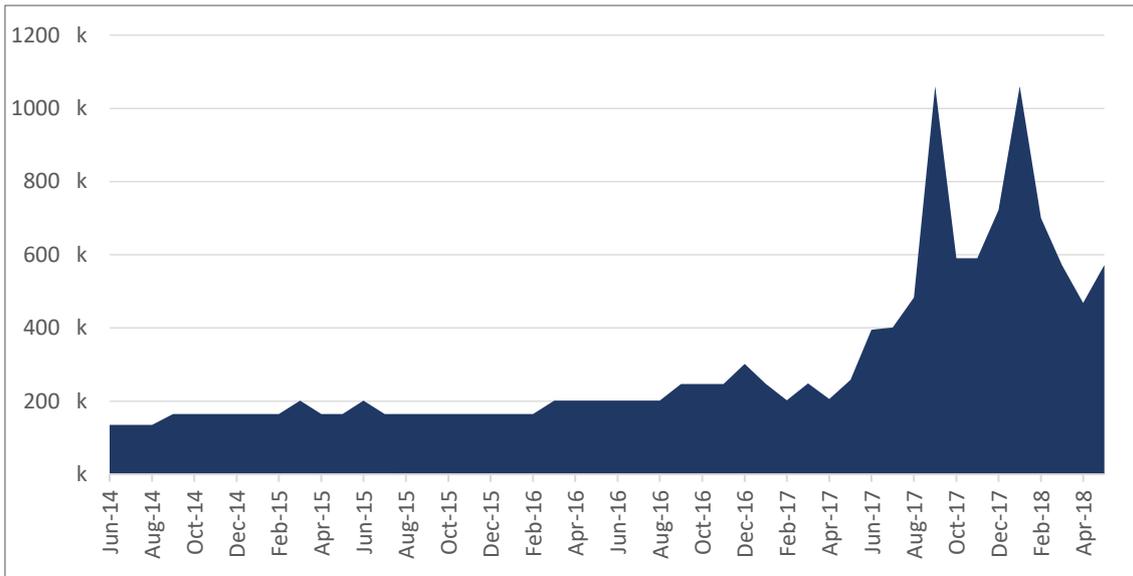
2016		2017		2018 (until May)	
Project	Total Raised	Project	Total Raised	Project	Total Raised
Waves	\$16,4 M	Hdac	\$258,0 M	Telegram ICO (Pre-sale 1 & 2)	\$1,700 B
Iconomi	\$10,6 M	Filecoin	\$257,0 M	Petro (Private Pre-Sale)	\$735,0 M
Golem	\$8,6 M	EOS Stage 1	\$185,0 M	Dragon	\$320,0 M
SingularDTV	\$7,5 M	Paragon	\$183,2 M	Huobi token	\$300,0 M
Lisk	\$5,7 M	Bancor	\$153,0 M	Bankera	\$150,9 M
Digix DAO	\$5,5 M	Status	\$90,0 M	Basis	\$133,0 M
FirstBlood	\$5,5 M	BANKEX	\$70,6 M	Orbs (Private Sale)	\$118,0 M
Synereo	\$4,7 M	TenX	\$64,0 M	Envion	\$100,0 M
Decent	\$4,2 M	Nebulas	\$60,0 M	Elastos	\$94,1 M
Antshares / NEO	\$3,6 M	MobileGO	\$53,1 M	U.CASH	\$84,9 M

Top 10 ICOs per year. Source: CoinSchedule

Evolution of the online demand⁸

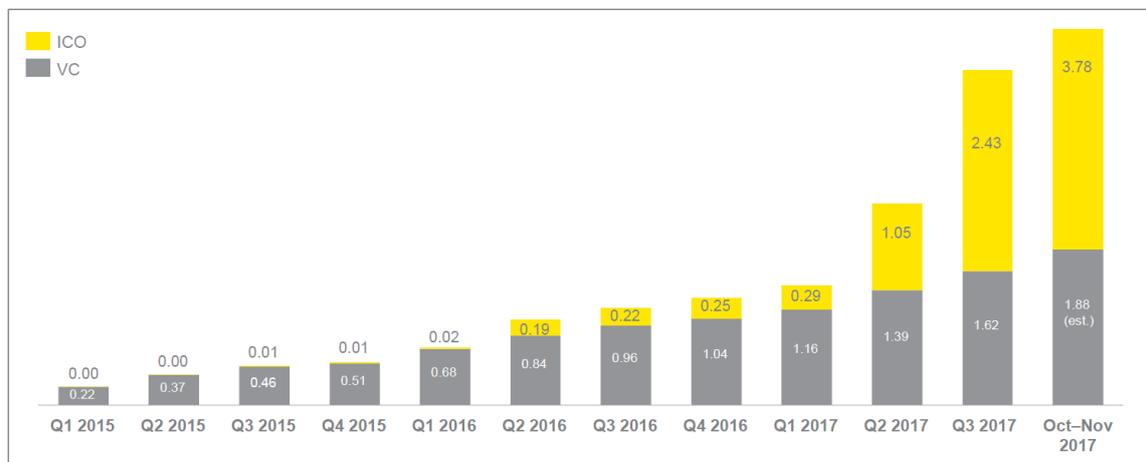
Online demand related to Initial Coin Offerings reached its highest peak in September 2017, with more than 1 million monthly Google searches³. Total annual searches went from 2,1 M in 2015 to 5,1 M in 2017, representing a CAGR of 41% in the period from June'14 to May'18. There was a massive increase of 109% between 2016 and 2017.

⁸ Source: Google AdWords



Evolution of the online demand of the keywords “ICO” and “Initial Coin Offering”. Source: Google AdWords

In terms of investment, during the 3rd quarter of 2017, total ICO proceeds exceeded venture capital investments in blockchain projects for the first time.



Evolution of ICO and VC funding in US\$ billion. Source: CoinDesk, CB Insights, IFCERT

4. CREATION OF AN ICO

The preparation and launching process of an ICO can be an even more important factor of success than the project itself. At some point, it has become a matter of marketing and social perception in the cryptocurrency community to generate

high expectations around it. A step by step of the process can be summarized as follows⁹:



Simplified diagram of an ICO process

a. Pre-announcement

Pre-announcement of the project on platforms such as Reddit or Bitcoin Talk to generate social traction among cryptocurrency investors. At the same time, the developing team publishes the white paper, a document that details the intentions of the project, its business model and the development roadmap, as well as the integration of the currency within its ecosystem.

This initial stage is also used to get a first impression of the generated investor interest and to get feedbacks and concerns about the project by would be investors to evolve and refine the business model and reach a final version of the white paper.

b. Offering

At this stage, the final version of the white paper includes a contract that outlines the project details, the terms of the transaction, the targeted invested amount and the lifecycle and deadlines of the project. It also details the value assigned to the token, the rights of the investors and the expected period to return earnings.

Once the offer entailed in the white paper is signed, the start date of the token sale is announced and the marketing campaign reaches its peak.

⁹ FXEmpire: "What is an ICO and how does it work?"

c. Marketing campaign

The current number of ICOs that are being launched daily (on average, more than 4,5 ICOs per day during May 2018¹⁰) has fostered the importance of an aggressive marketing campaign, determining at a great extent the success of any ICO. These campaigns, that cost tens of thousands of dollars and last up to a month on average, market the project in the cryptocurrency landscape in order to generate hype and social awareness to track investments before and during the token sale period.

d. Token Sale

The token sale begins at the date stated in the offering and despite it can be done in crypto exchanges, most ICOs enable a dedicated site within the web page of the project. Hence, these are the two main ways in which companies generally release tokens:

- Divide and distribute the tokens to the investors according to their initial investment after the specified amount of capital stated in the offer is collected
- Sell the tokens through cryptocurrency exchanges, having to release the tokens in advance on numerous exchanges for trading

Most of ICOs offer early-bird benefits in the form of pre-sales, with significant discounts on the token price (up to 25%) that decrease gradually in each successive pre-sale.

Despite some tokens can be purchased with fiat currencies, most ICOs prefer to be funded through cryptocurrencies, mostly Bitcoin and Ethereum.

¹⁰ Coinschedule

D. INVESTING IN ICOs

1. MAIN DIFFERENCES BETWEEN AN ICO AND AN IPO

The features of Initial Coin Offerings can be better understood when compared to Initial Public Offerings. An IPO is defined as the public sale of the shares of a company for the first time with the purpose of collecting funds for expansion and development. These shares, also called stocks, represent the perceived value of a company and are traded in stock markets.

So, while both practices share the same idea, there are some key differences between a cryptocurrency ICO and a stock IPO. Here are explained the main differences concerning regulation, record and credibility, utility and the nature of the offerings.

a. Regulation

Any company willing to issue an IPO must create a legal document called prospectus. The prospectus represents the legal intention to issue public shares and must meet strict standards of transparency, including key information about the company and the upcoming IPO as a form of information to potential investors.

On the other hand, ICOs are not required to issue any form of legal documentation at all. The developing team issues a white paper with the key information of the project including specifications about the ICO and the team itself. There are no standard procedures to create a white paper so its content is up to the team and its willing to persuade investors with relevant and transparent information.

b. Tracking record & credibility

Acting as a filter for credibility and trust, a company has to fulfil several requirements to be able to issue an IPO, such as a minimum earnings threshold and a good track record. It also requires professional accounting firms and

investment banks to provide verification and underwriting of the deal. Moreover, any real candidate of issuing and IPO has been previously funded by institutional investors that have already challenged the viability of the business through rigorous due diligences.

Given that ICOs are not regulated, most of them don't have any track record and the white paper and the developing team are the only elements to back up the project. In contrast with IPOs, where due diligences are focused towards historical performance, assessing an ICO is mainly based on future expectations and sensations around the experience of the team making of this process a very complex one.

c. Utility

Whereas the utility of holding stock issued is the entitlement of the shareholders in receiving dividends and voting in the shareholders meeting, acquiring tokens or coins in an ICO doesn't grant ownership of the project. In this case, future benefits of the token holders will depend on the structure of the coin and its perceived utility. While some coins will generate value conferring a stake in future revenues, others will be used within its ecosystem, increasing its value with its level of adoption.

d. Duration and access to the offerings

The issuance of an IPO can take up to 6 months due to the legal and compliance processes and the approval of regulatory authorities.

On the other hand, the duration of an ICO process is much shorter and depends on the nature and the timeline of the project. In fact, it only takes the whitepaper and a smart contract for the crowdsale, the length of which depends either on a fixed duration (typically one month) or on reaching a determined amount of funds raised.

While participation in IPOs is usually restricted to institutional investors like investment banks or funds, with small portions allocated to retail investors,

literally anyone can take part of an ICO through an easy and straightforward procedure that has democratized investing to the general public.

	ICO	IPO
Regulation	None	Comprehensive
Tracking record & credibility	Weak	Strong
Utility	Adoption	Dividends
Duration of offering	Short	Lengthy
Access to offerings	General public	Exclusive

Summary of the key differences between ICOs and IPOs

2. ADVANTAGES OF INVESTING IN ICOs

Investing in ICOs allows small individual investors interested in tech to indirectly participate and get involved in revolutionary projects that aim to disrupt the state of the art of the current industries and set new social and economic standards.

Unlike Initial Public Offerings, ICOs have no barriers of entry given that any investor can participate with relatively small amounts of capital and still reap huge returns in the event of success of the underlying project. Besides, initial investors can leverage the economic prominence of the supply-demand principle to increase the chances of exponential returns with the future gain of value of the token.

Linked to the mid and long-term success of the blockchain projects and its implementations into real practical applications, crypto-tokens can be used to purchase the services offered and in the case of high popularity and market trust, they can be also used for third-party purchases.

Moreover, one of the most interesting features of ICOs is that, compared to investing in traditional startups, where capital is often tied for long periods of time, tokens have liquidity at the moment they are traded in cryptocurrency exchanges, allowing investors to cash out and cash in at any time.

3. RISKS OF INVESTING IN ICOs

The counterparts of investing in ICOs are strongly linked with its benefits. The myriad of opportunities, the inexistent barriers of entry and the early-stage of the involved projects make them extremely attractive for small investors but, at the same time, these investments offer little or no financial guarantees and are subject to important uncertainties about the feasibility of the business model and the previous experience and background of the developing team.

According to a report by **Statis Group LLC**¹¹, released in early 2018, 80% of ICOs are scams, and only 8% manage to reach the trading stage on cryptocurrency exchanges. The study, performed on 2,300 ICOs with a minimum market capitalisation of \$50 M that were expected to be traded, categorized ICOs into six different groups based on its success. The results of the study are presented below:

Category	Group	Description	Distribution
Scam	Not listed	No apparent red flags that turned out to be scams	81%
Failed		Managed to raise capital but were abandoned at some point of the process and ended up refunding the investors' money	6%
Gone Dead		Funded as required through the standard process but wasn't listed on exchanges	5%
Dwindling	Listed	One or less success criteria*	4%
Promising		Two success criteria*	2%
Successfull		All success criteria*	2%

Success criteria

Deployment of a chain / distributed ledger for a base-layer protocol (test or beta version minimum)

Deployment of a product/platform for an app / utility token (test or beta version minimum)

Transparent project roadmap in the website

Contribution activity in the Github code during a period of three months

Results of the study performed by Statis Group LLC

¹¹ Statis Group LLC is a New York-based ICO advisory firm which also works as a digital asset focused investment bank <https://www.investopedia.com/news/80-icos-are-scams-report/>

Given the considerable amount of fraudulent ICOs, the following section aims to explain how to detect ICO scams.

a. How to spot an ICO scam

Below are listed the red flags to spot an ICO scam¹² grouped by the project and code repository, token analysis and project economics, team and community and whitepaper.

- **Project and code repository**
 - No fundamental scientific research on the white paper
 - No real necessity of a decentralized blockchain within the project
 - No code on public repositories such as Github and/or limited activity
 - Unrealistic goals without an economic plan or roadmap to support it
 - New cryptographic algorithm with limited feedback and/on algorithm controversy
- **Token analysis and project economics**
 - No long-term profitable economic model
 - Unclear about the need of a new token
 - Unclear about the value of the token (voting right, currency, stock...)
 - Founders award themselves with more than 50% of the tokens
- **Team & Community**
 - Lack of open team profiles
 - No backing by press and community
 - Very recent communication and social media accounts
 - Suspicious discussions on reddit or twitter
- **Whitepaper**
 - Starting by the coin sale without specifying the project enough

¹² HEC Paris: Finance at the digital era

- Too simple to understand, lacking technical details. Proper whitepapers detail their “disrupting” technology with pertinent charts, calculations, specifications and even code
- Multiple pre-sales periods with very high bonus discounts

4. HOW TO EVALUATE AN ICO

The best way to evaluate an ICO is to analyse every single aspect of the project and narrow it down to its precise execution strategy. Some critical factors to consider include:

- **Having a whitepaper:** A whitepaper is a public document, typically available on the project’s website, which describes the notion of the project. It is *sine qua non*¹³ for an ICO that the project is doomed to fail if the founders team do not have the capability to publish a proper whitepaper.
- **Business model:** Identify the strong points and weak links of the business model. Does the company really need an ICO with its own currency? How does the company integrate crypto token with its business model? Can you use the crypto token outside its ecosystem? Enquire from people who understand the technology and discuss these points with them.
- **Communication and Marketing:** It is essential to increase the peace of mind for the investors, as there have been many ICOs resulted in scams. Many so-called founders took millions from people and vanished. Means of communication is crucial to become transparent and enhance the performance of the ICO. Blog posts by opinion leaders, social media existence, newsletters to supporters, YouTube videos about the prototype or sharing LinkedIn profiles of founders can be effective ways of communication which can impact the success of the ICO. ICOs are announced on community forums and the marketing part kicks in afterward. Check the community feedback for the project and how the project team responds to it. Visit the Reddit, Twitter or Facebook pages of the

¹³ Means “Absolute Must” in Latin

projects. Be aware of paid reviews or bounty posts that pay participants to spread positive information about the project.

- **Timing:** It is important to launch the ICO, when the most optimum conditions are foreseen for the overall crypto market. Among the conditions which affect the crypto activity, one can consider the rise and fall of bitcoin, change of regulations, scandals, tax notifications, holiday seasons and forks.
- **Project team:** Find out every single detail you can about the development team and the advisory board. Look for ICOs that have teams with proven track record in the crypto-asset and blockchain industry.
- **Current product state:** ICOs that have a product near completion or halfway down the development track have higher chances. Find out the current project stage and whether the development team is on schedule or not.
- **VC involvement:** The involvement of a big crypto VC is a positive sign. Find out if the company has VC support or better if they have a VC on board.
- **Market niche:** How intense is the competition in the selected market niche? Research on any previous businesses working in the same industry, and match their fundamentals with what the company is working on.
- **Escrow agencies**¹⁴: investors can also look out for ICOs that include independent escrow agents, so that the capital raised does not reach the company entering an ICO, but a 3rd party. Multi-Sig is an example of such an escrow agent, with the agent essentially funding the project on an ongoing basis, funds released from an escrow as needed, the agent ensuring that project targets are being met along with the company's pre-determined obligations to the investor.

¹⁴ <https://medium.com/blockchain-hub/how-to-evaluate-a-ico-part-1-c6829d4de766>

5. THE DEFINITION OF SUCCESS FOR AN ICO

The dynamics of ICO vary. Some projects set a specific ICO period, some end the ICO after reaching their target leverage, some promote their tokens by giving bounties, As the dynamics of ICO vary, the expected return of investors from an ICO also varies. Basically, if the ICO manages to satisfy the expectation of an investor, the ICO is considered as successful. It is independent of investing an ICO which is scam or Ponzi. For example, The project can achieve to its fundraising goals that are predicted before the ICO, can be listed in the coin markets, but the founders might not have any interest in maturing the project. Still, there are investors who aim to become bullish and get short positions on the token that they can convert the crypto token into fiat currencies right after the token is listed and make a significant return. In this frame, this is a successful ICO for them, even if the project is doomed to fail in the longer term. In this paper, ICOs are clustered in two categories, namely *successful* and *unsuccessful*. We will try to predict the ICOs which will provide a return per token of more than 100% within the first 2 weeks, after the token is listed in the coin markets.

E. PREDICTING THE FINANCIAL SUCCESS OF ICOs BASED ON DATA AND THE WHITEPAPER ANALYSIS

1. CONTEXT

The whitepaper analysis and data analysis constitute the core of our research and predictions, as it gives a lot of hints about the likelihood of an ICO to succeed. In this paper, we are trying to predict what investors would be thinking about the future of an ICO based on their limited insights of the project and its founders. We realized that people are investing in ICOs because of the phenomenon called 'fear of missing out' instead of conducting a real valuation. The more people attempt to invest in an ICO the more its token circulates and increase its price per token. We also realized that people are investing after examining the whitepaper and based on their gut feelings about the project. That is why, the whitepaper analysis is crucial for modelling the success of the ICO. Below one can find a list of criteria which will be investigated in our whitepaper analysis.

- Prototyping an MVP
- Token distribution: Maximum supply is one of the most significant factors that affect the performance of price per token, the price of the token, and how much founders keep tokens and the funds generated for themselves
- Strategy
- Country of the establishment
- Roadmap: How the founders will be using the fund generated after the ICO, stages of the project
- Use cases of the token
- Building their own protocol
- Competitive advantage
- Partnerships

These criteria can be modelled by using text-mining algorithms. Basically, the context of the whitepaper is investigated by checking the terms inside the paper and how frequently they are used. After the text in each whitepaper is mathematically stored, they are categorised based on their success, which is ROI of ICOs, by using machine learning techniques, such K-nearest-neighbour (KNN)

algorithm. Then, the analysis is tested by creating a confusion matrix to reflect the accuracy of the modelling approach for the context of the whitepaper.

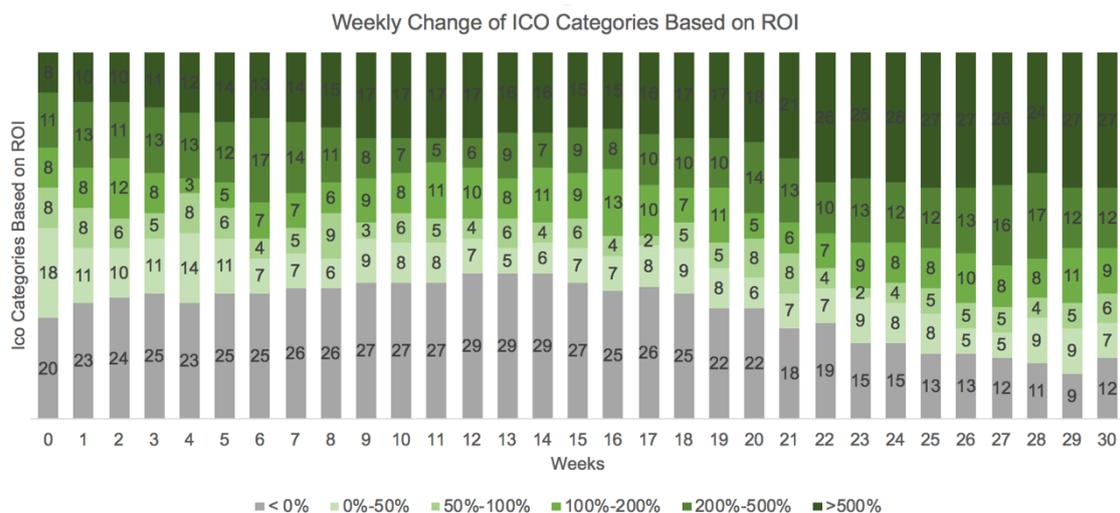
The following section hereby explains the ICO data extraction process to feed and train the algorithm, the analysis of this data and the resulting insights and conclusions and, finally, the text mining algorithm construction and results.

2. DATA COLLECTION

In order to feed the text mining algorithm, historical data from 106 ICOs was extracted from the CoinMarketCap database. The 1-year period of analysis included the ICOs that started from September 2016 onwards and ended before September 2017, in other words, from 1 Sept. 2016 to 31 Aug. 2017.

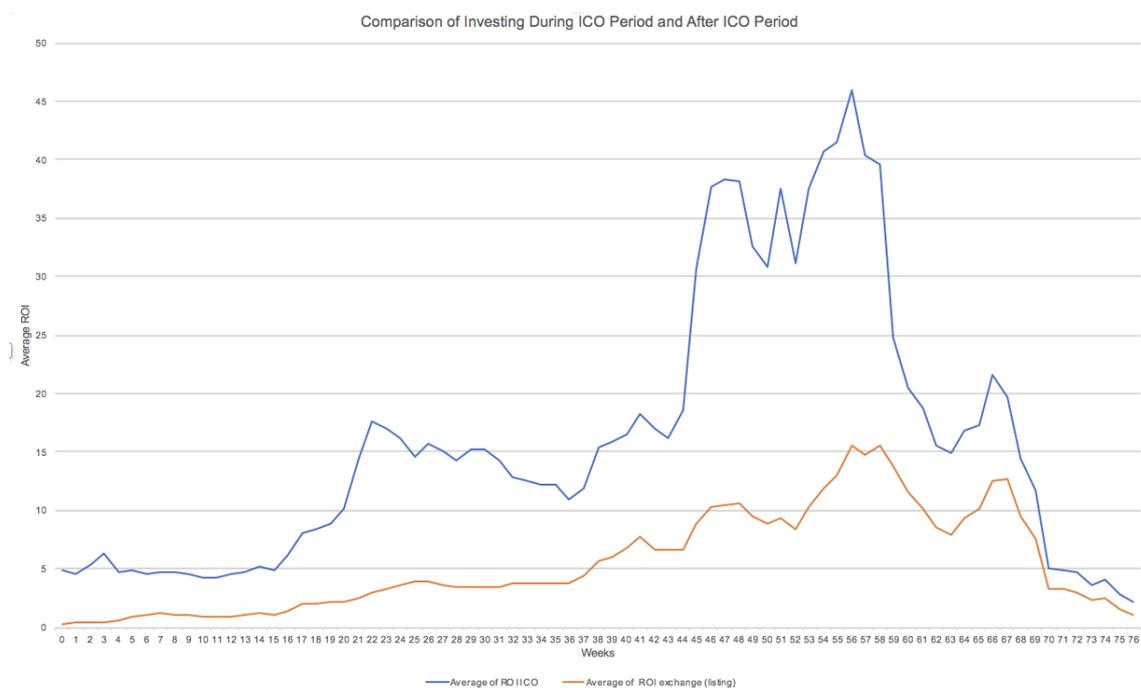
3. DATA ANALYSIS

First of all, the return of investment (ROI) is calculated based on the weekly average market price of 106 collected ICOs from “coinmarketcap.com”, as the main goal of our analysis was to understand the performance of ICOs from investors point of view. According to our analysis, the investors should wait for 27 weeks to exit from the project and convert their cryptocurrencies, as the graph below shows that 78% of ICOs bring ROI of more than 100% in week 27.

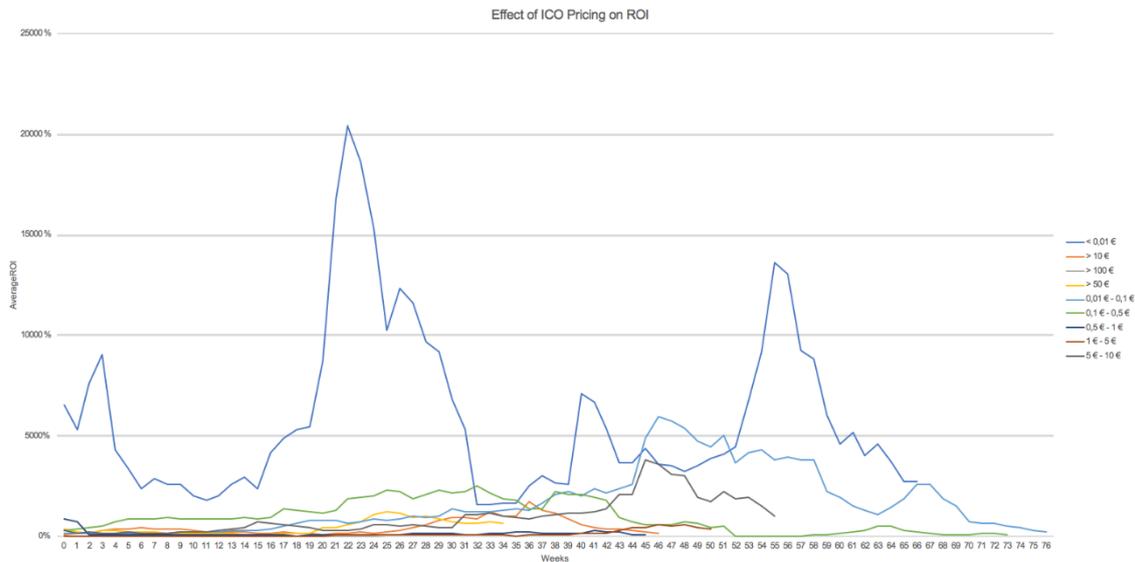


However, we also found out that only 70% of ICOs could survive 30 weeks and the rest fails. In other words, it is risky to wait for longer periods, but the return is more rewarding.

Furthermore, we analysed that investing during the ICO period for a cryptocurrency is more rewarding than investing in the same cryptocurrency after it is listed in coin markets, although there is always a risk that the project may become a scam. The graph below shows that investing during ICO is more profitable than investing after ICO.



Finally, we observed if the initial ICO price had an effect on ROI. The graph below shows that ICOs which are initially priced less than or equal to € 0.01 per coin are the ones which are most profitable.



We think that if the initial price were lower than € 0.01 per coin, investing in the corresponding coins would be more profitable. Statistical hypothesis testing is used to control our claim:

Average ROI for all ICOs is equal to 1104% for a sample of 91 ICOs and the average ROI for ICOs with initial price less than € 0.01 per coin is 6716%. In other words, our null hypothesis says that ICOs which are initially priced less than € 0.01 per coin are more profitable, as their ROIs are likely to be more than 1104%. We can use one-tailed test with 95% confidence interval (z-score=1.645) to check if our hypothesis is statistically significant.

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

<http://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/>

$$Z = (6.716 - 1.104) / (6.48 / (91^{0.5})) = 8.26$$

As 8.26 is bigger than the z-score for 95% confidence interval (1.645), It indicates that the lower initial coin price (price ≤ € 0.01) is statistically significant on ROI.

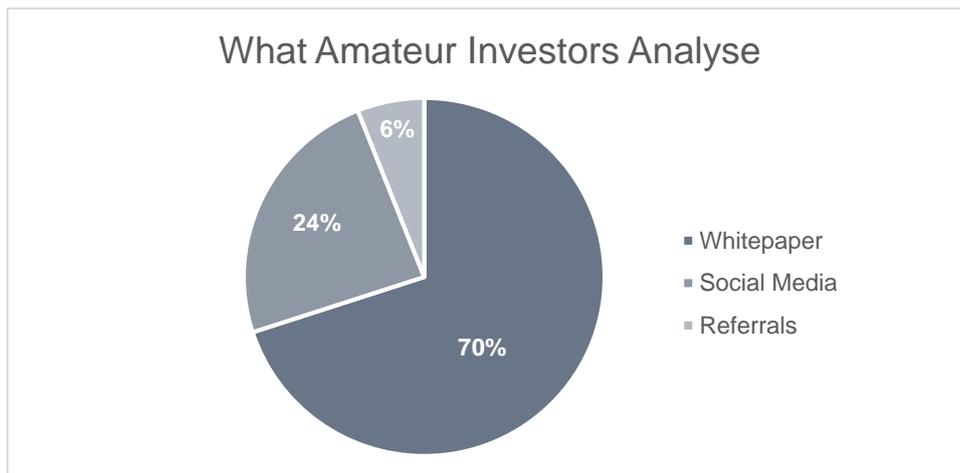
4. TEXT MINING ALGORITHM

As mentioned in previous sections, the success of an ICO is analysed under the scope of the return on investment throughout this paper. Basically, the goal is to find ways to automatize the analysis of the ICO projects and come up with a likelihood of achieving a significant return, after the ICO sale is over and the invested crypto currency is listed on the coin markets like Binance platform.

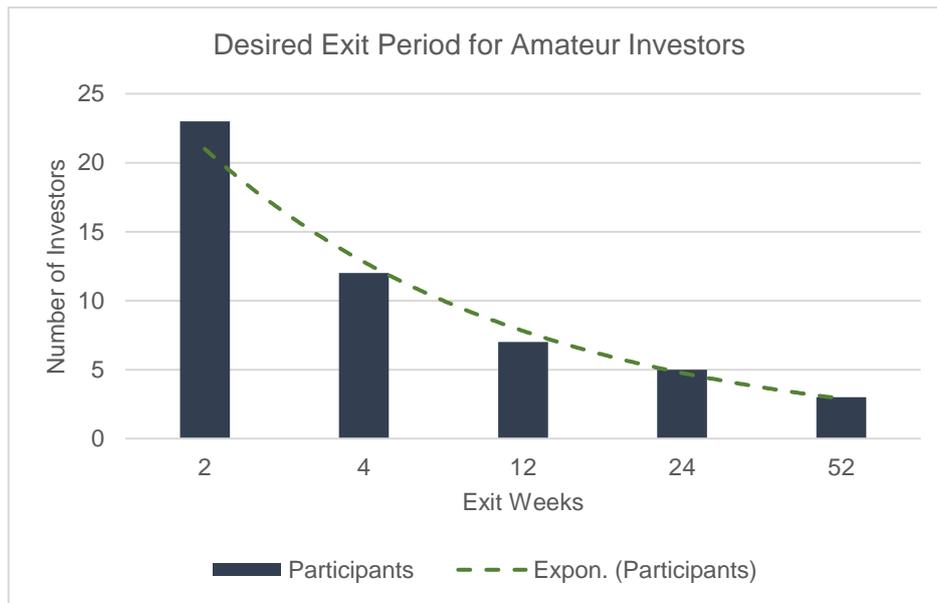
Automatization of analysis is a complex procedure for ICOs, as there is variety of data sources which are hard to extract such as marketing budgets of the ICOs, collaboration with the opinion leaders, the size of the community or social media influence, but give important insights for the sake of data. However, white papers of the ICOs have huge impact on the investors, especially the ones who are relatively amateur.

a. Survey to amateur investors

We conducted a survey among 50 amateur investors and the results indicated that the context of white papers has the most impact on the investment decision, which directed us to focus our automatization efforts on white paper analysis.

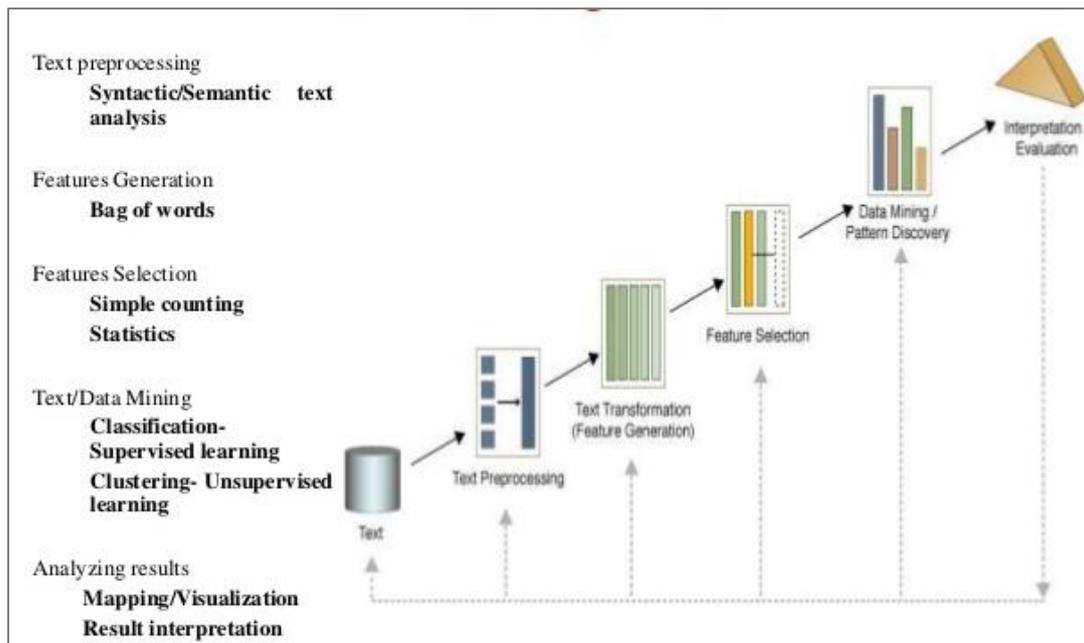


It is also understood that the main investment goal was to achieve a return of 100% or more within a 2-week period, after the ICO is listed on the coin markets.



Combining the 2 outcomes, our solution has been built on the classification of ICOs as *successful* and *unsuccessful*. After conducting some research on classifications, it is understood that supervised machine learning algorithms on text classifications, also known as text-mining, is the most appropriate technique for the analysis of white papers.

Text-mining is a way of smart classification technique by using natural language processing, which is one of the most powerful branches of supervised learning for real life applications. It is a process of deriving high quality information through statistical pattern learning. The notion of the text-mining is to process unstructured data and derive numerical matrices from the text to make the information ready for the machine learning algorithms. The entire process is properly summarized on the following figure.



Summary of text-mining process¹⁵

To elaborate more on the analysis, we would like to explain our algorithmic approach step by step:

1) Finding the dataset and the packages for text-mining for RStudio:

Over the courses of our analysis, we extracted pricing data for more than a hundred ICOs. We filtered them based on their return and grouped them in 2 clusters. The ICOs, which would provide a return of 100% or more within 2 weeks after being listed on the coin market for their investors are marked under the *successful* and the rest is marked as *unsuccessful*. Although, we collected data about more than a hundred ICOs, we could only find 59 proper whitepapers, 30 successful and 29 unsuccessful.

Open source RStudio packages, like *tm*, *NLP*, *plyr*, *datasets* and *class* are also benefited to facilitate the process of data mining.

2) Converting PDF documents to text documents:

White papers are published in the PDF format in general. However, it is harder to derive statistical information from PDFs, so they should have

¹⁵ <https://www.slideshare.net/kanimozhiu/text-data-mining1>

been converted into text files. The following piece of code would handle the issue:

```
#pdftotext

# Tell R what folder contains your PDFs

dest_successful <-
"/Users/macbook/Desktop/memoir/whitepapers_groups/successful"

dest_unsuccessful <-
"/Users/macbook/Desktop/memoir/whitepapers_groups/unsuccessful"

# make a vector of PDF file names

myfiles_successful <- list.files(path = dest_successful, pattern = "pdf",
full.names = TRUE)

myfiles_unsuccessful <- list.files(path = dest_unsuccessful, pattern = "pdf",
full.names = TRUE)

# PDF filenames can't have spaces in them for these operations

# so let's get rid of the spaces in the filenames

sapply(myfiles_successful, FUN = function(i){
  file.rename(from = i, to = paste0(dirname(i), "/", gsub(" ", "", basename(i))))
})

sapply(myfiles_unsuccessful, FUN = function(i){
  file.rename(from = i, to = paste0(dirname(i), "/", gsub(" ", "", basename(i))))
})

# get the PDF file names without spaces

myfiles_successful <- list.files(path = dest_successful, pattern = "pdf",
full.names = TRUE)
```

```

lapply(myfiles_successful, function(i)
system(paste("/Users/macbook/Desktop/memoir/xpdf-tools-mac-
4.00/bin64/pdftotext", paste0("", i, "")), wait = FALSE) )

myfiles_unsuccessful <- list.files(path = dest_unsuccessful, pattern = "pdf",
full.names = TRUE)

lapply(myfiles_unsuccessful, function(i)
system(paste("/Users/macbook/Desktop/memoir/xpdf-tools-mac-
4.00/bin64/pdftotext", paste0("", i, "")), wait = FALSE) )

```

3) Set Parameters:

Successful and unsuccessful tags are assigned as parameters and a pathname variable is created to access the whitepapers which are already downloaded on the computer.

```

#set parameters

candidates <- c("successful","unsuccessful")

pathname <- "/Users/macbook/Desktop/memoir/whitepapers_groups"

```

4) Cleaning the texts:

We need to figure out a way to create a model, where we can have some independent variables to train the machine learning classification model, to learn some correlations between the independent variables (the terms in the matrices) and the dependent variable (the categories). Now, the goal is simply to create independent variables. The idea is to create a model which would basically become a huge data table where the rows are white papers and the columns will be all the words that one can find in the white papers. Then, each cell will be indicating the frequency of words that appears in each white paper.

Since we will take all the words, we get a lot of columns. However, we don't want that scenario, because the more we have columns the harder it will be for the machine learning model to understand the correlations and

make accurate predictions. Some words give more hint to group the whitepapers and the task is to eliminate the words from the matrices which gives no clue about the category of the texts.

To sum up, we clean the punctuation, numbers, stopwords and white spaces from the corpus that we built from the text documents by using the following code:

```
#--clean the texts

#a corpus is simply a collection of documents

#create a corpus for each candidate and feed all the corresponding whitepapers

cleanCorpus <- function (corpus){ #just apply a variety of cleanup functions that are
available within tm

  corpus.tmp <- tm_map(corpus, removePunctuation)

  corpus.tmp <- tm_map(corpus, stripWhitespace)

  corpus.tmp <- tm_map(corpus, content_transformer(tolower))

  corpus.tmp <- tm_map(corpus, removeNumbers) # I added this

  corpus.tmp <- tm_map(corpus, removeWords, stopwords("english"))

  #stemming: getting the root of each word to decrease the number of tems.

  #e.g. handled and handle means same thing why would you have 2 terms

  #corpus.tmp <- tm_map(corpus, stemDocument)

  return(corpus.tmp)

}
```

5) Creating the Bag of Words Model:

We will create one column for each word and one row for each white paper in this section, which is known as Term Document Matrix Model. In other words, we are converting the text into a quantitative format that we can analyse from our corpus.

One of the most important things is to remove sparseness, while creating the bag of words model. A sparse matrix is a matrix that contains a lot of

zeros in its cells. Sparsity should be removed to filter infrequent terms in the matrix, so that the model can realize more accurate predictions.

It is done by the following piece of code:

```
#parameters are the name of the candidate(successful/unsuccessful) and the
path
generateTDM <- function (cand, path){
  #paper directory variable
  p.dir <- sprintf("%s/%s",path,cand)#print those 2 variables together
  #create a corpus based on path p.dir
  p.cor <- VCorpus(DirSource(directory = p.dir),readerControl = list(reader=
readPlain))
  #then apply the cleanCorpus function
  p.cor.cl <- cleanCorpus(p.cor)
  #after cleaning, create the tdm
  p.tdm <- TermDocumentMatrix(p.cor.cl)
  #A sparse matrix is a matrix that contains a lot of zeros in its cells.
  #Remove sparsity to filter infrequent terms in the matrix.
  p.tdm <- removeSparseTerms(p.tdm, 0.1) #0.1 stands for the parameter of
sparseness

  #return the result
  result <- list(name=cand, tdm=p.tdm)
}

#create a variable for the result
#loop through the 2 candidates and apply generateTDM function that I wrote
tdm <- lapply(candidates, generateTDM, path=pathname)
```

6) Attach the groups in the Term-Document-Matrix (TDM):

It is done for tagging whitepapers as successful or unsuccessful by using the following code:

```
#---take the tdm and attach the group (successful, unsuccessful)

bindCategoryToTDM <- function (tdm){

  #convert tdm into a numeric data matrix

  #and transpose the row and column variables to make it easier
```

```
#each whitepaper as a row and each term as a column (right now it's the other way around)
```

```
p.mat <- t(data.matrix(tdm[["tdm"]]))
```

```
#convert this matrix into a dataframe
```

```
p.df <- as.data.frame(p.mat, stringsAsFactors = FALSE)
```

```
# I wanna put the name of the success category to the last column
```

```
#use cbind() column bind to do that
```

```
p.df <- cbind(p.df, rep(tdm[["name"]], nrow(p.df))) #replicate the name by the number of rows in our matrix
```

```
colnames(p.df)[ncol(p.df)] <- "targetcategory"
```

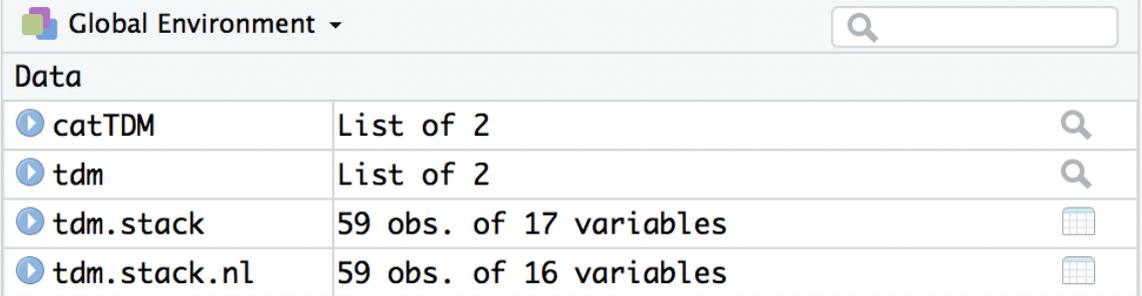
```
return(p.df)
```

```
}
```

```
catTDM <- lapply(tdm, bindCategoryToTDM)
```

7) Stack the 2 matrices:

To mention about our dataset, we have 59 rows which represent the white paper names and 17 independent variables which contain the most frequent terms and help model to realize the most accurate predictions.



Global Environment		
Data		
▶ catTDM	List of 2	🔍
▶ tdm	List of 2	🔍
▶ tdm.stack	59 obs. of 17 variables	📅
▶ tdm.stack.nl	59 obs. of 16 variables	📅

Dataset which shows that we have a list of 2 categories (successful and unsuccessful), 59 white papers and 16 variables which are the most frequent terms in the Term Document Matrix

To show it in a simpler way in order to facilitate the analysis, we stack successful and unsuccessful ICOs one on top of the other by using the following code:

```
#---take each of the 2 matrices and stack them one on top of the other  
  
#we're gonna feed them all in one time into our model  
  
tdm.stack <- do.call(rbind.fill, catTDM)  
  
tdm.stack[is.na(tdm.stack)] <- 0 #NAs are set to 0
```

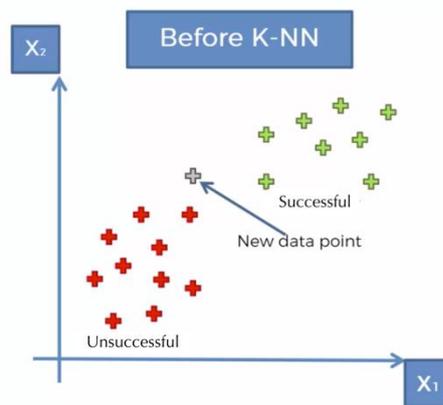
8) Create a hold-out sample:

We take a portion of our data to train or build the model and we also hold a small subset from our dataset, which is referred as sample set to let our model the make predictions. More specifically, we use 70% of our 59 white papers to train the model and the remaining 30% is extracted as a sample set to make predictions.

This is done by the following code:

```
#---create a hold-out sample  
  
#we take a portion of the whitepapers and we use them to train/build the model  
  
#a small subset of the whitepapers, we wont tell which whitepaper is that  
  
#and we will let the model to predict  
  
#these are going to be the rows which teach to model  
  
#take 70%  
  
train.idx <- sample(nrow(tdm.stack),ceiling(nrow(tdm.stack)*0.7))  
  
test.idx <- (1:nrow(tdm.stack)) [-train.idx]
```

9) K Nearest Neighbour (KNN) Model:



KNN description for representing our dataset

In our dataset, 2 categories are identified, successful and unsuccessful. Each dot represents a whitepaper. We have shown only variables x_1 and x_2 for the sake of simplicity. The figure shows that the classification is based on x_1 and x_2 but we have actually 17 variables (columns) which determines our model in our final TDM matrix.

The purpose of the model is to decide the classification of a new data group. In other words, to decide if it should fall into the successful category or unsuccessful category.

The main issue is how to decide it. That's where KNN helps us. The first step is to pick number K for neighbours. One of the most common default values for K is 5. Then, the Euclidean distance to each data point is calculated and K nearest neighbours are selected. Among these K neighbours, you need to count the number of neighbours in each category. Finally, the new data point is assigned to the category where you counted the most neighbours.

The process is handled by the following piece of code:

```
#---model KNN

#include all rows but just the success category as column

tdm.category <- tdm.stack["targetcategory"]

#all of the columns except the success category
```

```
tdm.stack.nl <- tdm.stack[, !colnames(tdm.stack) %in% "targetcategory"]
```

```
#for knn model: plugin the training set and the test set, but just give names for  
the training set for learning
```

```
knn.prediction <- knn(tdm.stack.nl[train.idx, ], tdm.stack.nl[test.idx, ],  
tdm.category[train.idx])
```

10) Assess the accuracy:

One way of checking the accuracy is setting up a confusion matrix, which is a specific table layout to visualize the performance of a supervised learning algorithm. Basically, it shows how our model predicted on the test set against texts that it didn't know how the classification was. It compares the predictions with the actual results that we know how the white papers should have been classified.

The following code is used for creating the confusion matrix:

```
#---assess the accuracy  
  
#set the confusion matrix (shows how our model predicted on the test set)  
  
conf.mat <- table("Predictions" = knn.prediction, Actual = tdm.category[test.idx  
)  
  
  
accuracy <- sum(diag(conf.mat))/length(test.idx) * 100
```

The result of the accuracy of our model is between 95% and 100%, which is a very strong result and shows how relevant our approach is to automatize the analysis of whitepapers by implementing text-mining algorithms.

```
> conf.mat  
      Actual  
Predictions successful unsuccessful  
successful      10             0  
unsuccessful     0             7  
> View(tdm.stack.nl)  
> View(tdm.stack.nl)  
> accuracy  
[1] 100  
> |
```

Print of the confusion matrix and the accuracy of the model

The figure shows the confusion matrix and the accuracy of the model which is 100%. Our model made predictions on 17 white papers in the test set, which amounts to 30% of the 59 white papers which we included to our TDM in total. On the matrix, it is observed that the model does not make any mistake on predicting 10 successful ICOs and 7 unsuccessful ICOs.

F. CONCLUSION

“Blockchain refers to a distributed, encrypted database, which is a public depository of information that cannot be reversed and is incorruptible.”¹⁶ This technology is a nascent, promising technology which can have an impact on transforming businesses, increasing the prosperity of the society and shaping the future. Its potential to decentralize the control of powerful institutions and to disrupt the traditional businesses makes the technology face potential challenges such as regulative issues. However, it is very likely that people will be embracing Blockchain, as they gradually discover and understand how it is the use cases of the Blockchain technology can help boost the trust mechanism between parties, increase the efficiency of a business and augment the prosperity of the society.

Executives and visionaries have already explored many use cases of blockchain. However, the discovery of using cryptocurrencies, which are created by using blockchain, as means of crowdfunding in the form of initial coin offering (ICO) had a huge impact on shaping the way people invest in 2017.

The ICO market boomed in 2017, with 210 ICOs raising a total of \$3.9 B¹⁷, a figure 40 times higher than the one raised in 2016, when 43 ICOs were created. Far from being a transient trend, the volume of ICO investment keeps growing exponentially, with \$9,8 B raised so far in 2018 and 415 ICOs created.¹⁸

Investing in ICOs allows small individual investors interested in tech to indirectly participate and get involved in revolutionary projects that aim to disrupt the state of the art of the current industries and set new social and economic standards.

Linked to the mid and long-term success of the blockchain projects and its implementations into real practical applications, crypto-tokens can be used to purchase the services offered and in the case of high popularity and market trust, they can be also used for third-party purchases.

¹⁶ Morabito, V.: Business Innovation Through Blockchain

¹⁷ Valuation based on BTC exchange rate at the time of the ICO closing date. Source: CoinSchedule

¹⁸ Source: CoinSchedule

The counterparts of investing in ICOs are strongly linked with its benefits. The myriad of opportunities, the inexistent barriers of entry and the early-stage of the involved projects make them extremely attractive for small investors but, at the same time, these investments offer little or no financial guarantees and are subject to important uncertainties about the feasibility of the business model and the previous experience and background of the developing team.

According to a report by **Statis Group LLC**¹⁹, released in early 2018, 80% of ICOs are scams, and only 8% manage to reach the trading stage on cryptocurrency exchanges.

The best way to evaluate an ICO is to analyse every single aspect of the project and narrow it down to its precise execution strategy. Some critical factors to consider include whitepaper, business model, communication and marketing, current state of project, project team and VC involvement.

The dynamics of ICO vary. Some projects set a specific ICO period, some end the ICO after reaching their target leverage, some promote their tokens by giving bounties, As the dynamics of ICO vary, the expected return of investors from an ICO also varies. Basically, if the ICO manages to satisfy the expectation of an investor, the ICO is considered as successful.

In this paper, ICOs are clustered in two categories, namely *successful* and *unsuccessful*. We will try to predict the ICOs which will provide a return per token of more than 100% within the first 2 weeks, after the token is listed in the coin markets.

Both the whitepaper analysis and the data analysis constitute the core of our research and predictions, as it gives a lot of hints about the likelihood of an ICO to succeed. Basically, the context of the whitepaper is investigated by checking the terms inside the paper and how frequently they are used. After the text in each whitepaper is mathematically stored, they are categorised based on their

¹⁹ Statis Group LLC is a New York-based ICO advisory firm which also works as a digital asset focused investment bank <https://www.investopedia.com/news/80-icos-are-scams-report/>

success, which is ROI of ICOs, by using machine learning techniques, such K-nearest-neighbour (KNN) algorithm. Then, the analysis is tested by creating a confusion matrix to reflect the accuracy of the modelling approach for the context of the whitepaper.

In order to feed the text mining algorithm, historical data from 106 ICOs was extracted from the CoinMarketCap database. The 1-year period of analysis included the ICOs that started from September 2016 onwards and ended before September 2017.

The return of investment (ROI) is calculated based on the weekly average market price of 106 collected ICOs from “coinmarketcap.com”, as the main goal of our analysis was to understand the performance of ICOs from investors point of view. According to our analysis, the investors should wait for 27 weeks to exit from the project and convert their cryptocurrencies. Furthermore, we analysed that investing during the ICO period for a cryptocurrency is more rewarding than investing in the same cryptocurrency after it is listed in coin markets. In addition, we thought that if the initial price were lower than € 0.01 per coin, investing in the corresponding coins would be more profitable. One-tailed test is used to control our hypothesis. The test indicated that the lower initial coin price was statistically significant on ROI.

Whitepapers have huge impact on the investors, especially the ones who are relatively amateur. We conducted a survey among 50 amateur investors and the results indicated that the context of white papers has the most impact on the investment decision, which directed us to focus our automatization efforts on white paper analysis. It is also understood that the main investment goal was to achieve a return of 100% or more within a 2-week period, after the ICO is listed on the coin markets. After conducting some research on classifications, it is understood that supervised machine learning algorithms on text classifications, also known as text-mining, is the most appropriate technique for the analysis of white papers. It is a process of deriving high quality information through statistical pattern learning. The notion of the text-mining is to process unstructured data and derive numerical matrices from the text to make the information ready for the

machine learning algorithms. Our text-mining model, which uses KNN algorithm for machine learning, made predictions on 17 white papers in the test set, which amounts to 30% of the 59 white papers which we included to our term-document-matrix in total. By creating a confusion matrix, it is observed that the model does not make any mistake on predicting 10 successful ICOs and 7 unsuccessful ICOs.

To sum up, ICOs have become one of the first use cases of blockchain that changed a highly regulated industry as is finance. They have been in high demand since the success of Ethereum Smart Contracts. However, many impostors blossomed to benefit from the enormous demand, especially over the course of 2017. In this paper, we put forward the points to be considered before investing in ICOs and supported our assumptions by conducting data analysis on 106 ICOs and surveys on 50 amateur investors. Based on our study, we realized that one of the major pain points is to analyse white papers, as it is both time consuming and amateur investors struggle to understand the maturity, the potential use cases and the value proposition of the projects. In this context, we developed a text-mining algorithm with accuracy between 95%-100% by using R Studio and its natural language processing libraries to apply machine learning techniques on white papers in order to automatize the investment decision based on the scope of the project.