J. Int. Financ. Markets Inst. Money 83 (2023) 101716

Contents lists available at ScienceDirect

Journal of International Financial Markets, Institutions & Money

journal homepage: www.elsevier.com/locate/intfin

Fear sells: On the sentiment deceptions and fundraising success of initial coin offerings☆

Niranjan Sapkota*, Klaus Grobys

University of Vaasa, Vaasa, Finland

ARTICLE INFO

Keywords:
Initial Coin Offering (ICO)
Natural Language Processing (NLP)
Sentiment Dictionaries
Deep Learning
Artificial Neural Networks (ANN)

ABSTRACT

Retrieving information from an intensive hand-collected whitepaper data library covering 5,033 ICOs launched before 2020, we analyze the determinants of ICO success as measured by the amount of raised funding. We assess the sentiment and readability in ICO whitepapers in addition to other information disclosures. Whereas we do not find any evidence for that the riskiness of ICO projects would lower the predicted amount of raised funding, our results strongly suggest that ICO investors are largely guided by emotions when making investment decisions. Contrary to earlier literature, we find a weak association between quality signals of whitepapers and its success.

1. Introduction

Recently, initial coin offerings (ICOs) have received considerable attention as a new form of crowdfunding based on blockchain technology. Recent research documents that more than $30 billion has been raised via the ICO market (Howell, Niessner, and Yermack,

☆ We gratefully acknowledge the Project Research Grant (Grant no. 190405) by the Foundation for Economic Education (LIIKESIVISTYSRAHASTO), Finland. We would like to thank the discussant Professor Gonul Colak, and other participants at the Summer Workshop 2021 for providing us insightful comments to our paper (the Summer Workshop 2021 was organized by Aalto University in association with the Graduate School of Finance (GSF), Helsinki, Finland). We are also grateful to the discussant and the participants of the World Finance Conference (University of Agder, Norway (Virtual), August 3-6, 2021) for providing us useful and encouraging comments. We also received useful comments from the participants of the Cryptocurrency Research Conference 2021 (September 16-17, 2021, organized by the University of Southampton, UK and ICMA Centre, UK). Moreover, we also received insightful and encouraging comments from the participants of the IRMC 2021 (14th Edition October 1-2, 2021, organized by the Risk Banking and Finance Society, in collaboration with the University of Florence, NYU Stern Salomon Center, USA along with the co-organizer Joint Research Centre (JRC)-European Commission and the host institution, the University of Cagliari, Italy). Furthermore, we are thankful to Professor John Paul Broussard University of Oklahoma, Price College of Business, Finance and Professor John Goodell University of Akron, College of Business, Finance for providing us useful feedback.

* Corresponding author at: School of Accounting and Finance, University of Vaasa, Wolffintie 34, 65200 Vaasa, Finland
E-mail addresses: niranjan.sapkota@uva.fi (N. Sapkota), klaus.grobys@uva.fi (K. Grobys).

https://doi.org/10.1016/j.intfin.2022.101716
Received 6 April 2022; Accepted 16 December 2022
Available online 21 December 2022

2020). Due to their nature as unregulated offerings of digital tokens on the Internet, aiming to collect funding for a project, ICOs disintermediate any external platform, payment agent or professional investor and thus disrupt the current financial system, i.e. the market for Initial Public Offerings (IPO).[1] Unsurprisingly, due to its easy-to-execute approach to attaining external funding, ICOs have recently attracted enormous attention.

Recent finance literature on ICOs explores the link between ICOs price responses and investor attention (Tsukioka, Yanagi, and Takada, 2018), potential factors affecting ICO market outcomes (Lyandres et al., 2022; Zhang et al., 2022; Momtaz, 2020; Domingo, Piñeiro-Chousa and López-Cabarcos, 2020; Yu, 2019), the link between determinants of the characteristics of the advisory board and ICO fundraising success (Giudici, Moncayo, and Martinazzi, 2020), the usefulness of information availability as a market signal of quality (Meyer and Ante, 2020; Amsden and Schweizer, 2018), and the link between Twitter followers and activity and the success of ICOs (Benedetti and Kostovetsky, 2021). Other important contributions are the studies of Roosenboom, van der Kolk, and de Jong (2020) and Howell, Niessner, and Yermack (2020) documenting that ICO success is associated with (*i*) disclosure, (*ii*) credible commitment to the project and (*iii*) quality signals such as token listings. Finally, the study which is perhaps the closest related to ours is the one of Fisch (2019) that uses data on 431 ICOs to investigate the role of signalling ventures' technological capabilities in ICOs. Fisch's (2019) results indicate that technical whitepapers and high-quality source codes increase the amount of raised funding.

Motivated by this recent stream of literature, the purpose of this study is to explore the factors that determine the success of ICOs in terms of raised funding. In doing so, a novel issue that we consider is whether sentiment embedded in ICO whitepapers serves as a predictor variable for ICO success. To do so, we retrieve a unique hand-collected data set using all 5,033 ICOs that were launched between August 2014 to December 2019. We combine and match the information from ICO whitepapers with various databases allowing us to identify plenty of ICO-specific information. In exploring the sentiment embedded in whitepapers, we applied four different sentiment dictionaries in association with seven different readability measures.

Our study contributes to the recent literature in various fundamentally important aspects. First and foremost, taking the broader finance perspective, our paper adds to the literature on entrepreneurial finance (e.g., Catalini and Gans, 2018; Chod and Lyandres, 2021; Kaal, 2018; Huang, Meoli, and Vismara, 2020; Lyandres, Palazzo, and Rabetti, 2019; Li and Mann, 2018). Specifically, our study adds to the literature exploring the determinants of success in ICOs (Adhami, Giudici, and Martinazzi, 2018; Fisch, 2019; Howell et al., 2020) by first (*i*) accessing the entire population of ICOs, that is, we retrieve all 5,033 ICOs launched in August 2014 to December 2019 period. As a consequence, our study is not exposed to potential small sample biases as it accounts for the whole population of available data. Second, (*ii*) we employ a total of 36 potential predictor variables that could have an impact on the success of ICOs. In doing so, we replicate Fisch's (2019) method as the raised amount represents our dependent variable. Given the current research context, there is no other study available covering our extensive data set and extracting such a large set of potential predictor variables. Moreover, Fisch (2019) uses ICO whitepapers for extracting the word count characteristics only. Our study does not only extract the word count characteristics of the ICO whitepapers but also other important characteristics such as sentiments, emotions and readability. Besides the ICO whitepaper, our study also incorporates, the possibility of an ICO project ending up as a scam, using 'Risk Score' as one of the predictor variables. On top of that our statistical model also accounts for social media hype, measured as the "Hype Score." Whereas Fisch (2019) accounts for qualitative disclosure in terms of high-quality source codes, patents and copyright, our model incorporates qualitative disclosures such as Country of Origin disclosure, Roadmap/Milestone disclosure, etc. and quantitative disclosures such as SoftCap, HardCap, Number of Tokens, Number of Categories, Team Size, etc.

Next, the psychological literature has repetitively confirmed the priority of processing emotion words against neutral words. (Anooshian and Hertel, 1994; Chen, Lin, Chen, Lu, and Guo, 2015; Kissler, Herbert, Peyk, Junghofer, 2007; Yap and Seow, 2014; Zhang et al., 2014; Zhao et al., 2018). In this regard, researchers in psycholinguistics and linguistics distinguished two kinds of emotion words, that is, "emotion words" and "emotion-laden words" (Altarriba, 2006; Pavlenko, 2008). Wu, Zhang and Yuan (2021) argue that behavioral and electrophysiological studies supported that also emotion-laden words (e.g., war, death, disaster, risk) affect human behavior. The psychological literature has not yet explored the effect of emotion-laden words in whitepapers. Due to the enormous amount of money involved in the market for ICOs, this is definitely not a trivial issue which needs to be investigated. In this regard, our study is the first that explores the psychological sentiment in ICO whitepapers and clarifies the following questions: Firstly (*i*), which emotional content dominates whitepapers, and secondly (*ii*), how psychological sentiment affects the success of ICOs.

Another important novel aspect of our study is that we explore the question of whether financial sentiment–as opposed to psychological sentiment–cached in whitepapers has an impact on the success of ICOs.[2] There are various sentiment dictionaries from

[1] In corporate finance, IPO has several requirements, such as a good track record of earnings above a minimum earnings threshold, whereas other financial criteria are set by the exchange where the firm plans are listed. Whereas the NASDAQ requires a total of $11 million pre-tax earnings in the previous three years and more than $2.2 million in each of the two most recent years, none of these financial requirements apply for ICOs. Generally, anyone who has an innovative idea or is willing to create a company, is eligible to issue an ICO. One could even argue that companies that are financially qualified for an IPO are overqualified for an ICO. Furthermore, firms might have to wait for many years before fulfilling the criteria set by the stock exchanges to issue an IPO. In this regard, using CRSP data, Jovanovic and Rousseau (2001) show that it takes years and even decades for firms to be listed on the stock exchange.

[2] In general, sentiment is a genre of the appraisal theory. Specifically, sentiment analysis investigates opinions expressed in texts and comprises (*i*) the extraction of opinion polarity (positive or negative), (*ii*) the target (or specific aspects of the target) to which the opinion refers, (*iii*) the holder of the opinion, and (*iv*) the time at which the opinion was expressed (Muhammad, Wiratunga, and Lothian, 2016). Indeed, the saying *It is not what you say that matters but the manner in which you say it; there lies the secret of the ages* (William Carlos Williams) indicates that the tone of a text is perhaps more influential than its substantive content. In fact, plenty of studies have been devoted to exploring the sentiment of news content, political speeches, and advertisements.

which sentiment scores can be calculated. Studies comparing sentiment in the finance-specific domain have focused on Henry's (2008) and Loughran and McDonald's (2011) sentiment dictionaries. Using the Harvard-IV general dictionary, Loughran and McDonald (2011) found its word list to be largely inapplicable to financial contexts and created a finance-specific list. Henry (2008) captured the tone of earnings press releases to create a word list for financial texts. These studies found a finance-specific dictionary to be more powerful than the general psychological dictionary (Sapkota, 2022). There are already many studies (for example; Li et al., 2014; Alessia et al., 2015; Pröllochs, Feuerriegel, and Neumann, 2015; Yekrangi and Abdolvand, 2021) on finance domain-specific sentiments as opposed to psychological sentiment. This is because the previous studies have shown that the borrowed sentiment dictionary from a different discipline is likely to misjudge the true sentiment in that particular context.

A growing body of research suggests that affect is a central component of individual decision-making, political judgment, and especially the processing of media contents. Notably, Young and Soroka (2012) highlight that negative affect seems to be extraordinarily important in the human psyche, and in political interactions. The whitepaper of an ICO is of major importance as it reveals the intended production outcome of the proposed business project: Consequently, potential investors may or may not invest in the ICO merely based on its content. Hence, several natural and important questions arise: First, (*i*) does it matter how an ICO whitepaper is conducted with respect to its content? Second, (*ii*) should the whitepaper be written in simple terms for easy readability so that even a naïve investor is able to grasp the project idea? Third, (*iii*) should the whitepaper's choice of words trigger the positive or the negative side of the sentiment in order to successfully attract investors? While earlier studies explored sentiment associated with IPOs (Loughran and McDonald, 2013; Bajo and Raimondo, 2017; Guldiken, Tupper, Nair, and Yu, 2017), our study is the first that seeks to answer these three important sentiment-related questions for ICO whitepapers.[3]

A final contribution of our study is a methodological one. Specifically, our study uses a different approach than previous studies. Earlier studies mainly focus on the performance of the ICO (e.g. token performance as measured in terms of returns), which is an ex-post ICO phenomenon (for example, Howell et al., 2020). Our paper extends earlier research by first (*i*) focusing on the sentiment side of the ICO and second, (*ii*) by utilizing cross-sectional data to identify potential key factors in determining the size of the raised amount. It is important to note that in doing so, our study controls for a battery of factors evidently associated with ICO success.

Identifying a total of 36 potential predictor variables, our result shows that investors in Financial Technology (FinTech) sectors are not immune to behavioral biases. The key results of our study can be summarized as follows. First, only the Harvard Psychological Sentiment Dictionary appears to provide useful information that can be linked to ICO success. Specifically, negative sentiment is associated with a higher amount of raised funding, whereas positive sentiment does not have any significant impact. Our study identifies that the prevalent emotion cached in whitepapers is 'fear'. Factorizing this emotion into its specific components shows that investors' behavior in the ICO market is mainly driven by fears associated with 'risk', 'problem', 'change', and 'regulation', among others.

Next, another important finding is that popularity in terms of media attention is a key determinant for the success of ICOs. We observe a linear relation as we move from a low to a high level of media attention. Whereas a low level of media attention does not correlate with ICO success, a higher level of media attention is associated with an increase in the amount of raised funding. Our results support the study of Benedetti and Kostovetsky (2021) in recognizing the impact of Twitter: Specifically, our findings indicate that a higher number of followers on Twitter correlates positively with ICO success. In this regard, a novel finding of our study is that signature campaigns are of significance. Signature campaigns–often referred to as bounty programs–may have different procedures[4].

An unforeseen finding is that readability does not have any impact on the success of ICOs, which is in stark contrast to what has been documented in the corresponding literature on IPOs (Loughran and McDonald, 2013; Bajo and Raimondo, 2017; Guldiken, Tupper, Nair, and Yu, 2017).[5] Further, team size only marginally influences ICO success, whereas risk assessments, country disclosure, category (e.g., industry), or the similarity of a whitepaper with another project's whitepaper do not. Finally, the evidence documented in the current research suggests that ICO investors are, generally, not acting as rational investors because they (*i*) are biased towards negative sentiment, (*ii*) do not take into account the risk assessments, and (*iii*) do not even consider whether a whitepaper is conducted in an understandable manner or (*iv*) if it violates copyrights.

This paper is organized as follows: The next section provides a literature review. The third section presents the data and methodology. Furthermore, the fourth section documents the results and the last section concludes.

2. Literature Review

In the wake of the increase of social media and blogging websites, textual sentiment analysis has increased significantly. Nowadays, companies are using Twitter and Facebook to analyze the sentiment of their clients. Since user-generated content such as posts, shares, likes, tweets and retweets are openly available, firms and companies have enormous opportunities to understand the customers (He

[3] Apart from these first-order questions that are specifically related to the ICO whitepaper, there are also second-order questions that arise. For instance, what about other characteristics of ICOs that are usually not found in whitepapers, such as social media followers (e.g., as measured in terms of *Hype Score*), projects backed up by people disclosing their identity (e.g., as measured by Know Your Customers *KYC Score*), or potential risk for fraud (e.g., as measured by *Risk Scores*)? Do these factors also have an impact in attracting potential investors?

[4] Essentially, a signature campaign is a subscription campaign where ICOs release signatures with an embedded code. The bounty stake associated with the campaign is based on the ranking of the participants. Generally, for most bounty campaigns, only people on Bitcointalk forum who are at least ranked as Junior Member can participate. We find that signature campaigns appear to be useful predictor variables for ICO success.

[5] This finding is also contrary to Fish's (2019) study, which finds that the way a whitepaper is conducted has an impact on ICO success.

et al., 2015). With the rise of digital crowdfunding, there are tremendous opportunities for the clients to understand their companies too. Openly available ICO whitepapers consist of essential information about the startups. As a consequence, potential investors have the opportunity to assess the quality of the project by understanding the hidden sentiments in the whitepaper. Investors can also perform the sentiment analysis of any social media or the blog posts shared by the startups. Previous studies have identified various factors associated with the success of ICOs and in this study, we group these papers into four different clusters.

2.1. Information disclosure and ICO success

Howell et al. (2020) find that the success of ICO depends on the disclosure, credible commitment to the project along with other quality signals. Their study shows that ICO token exchange listing causes higher future employment and giving access to token liquidity has a positive outcome for the enterprise. Using a database of 1,000 ICO whitepapers, Zetzsche et al. (2019) show that many ICOs offer inadequate disclosure of information: the majority of the ICO whitepapers are either silent on the initiators or backers/ promotors or do not provide contact details. Furthermore, more than half of the ICOs do not elaborate on the applicable law, segregation or pooling of client funds, and the existence of an external auditor[6]. Therefore, the decision to frequently invest in ICOs can perhaps not be the outcome of a rational thought process. Similarly, using hand-collected data on 472 public token sales over the period of 2013–2017, Boreiko and Risteski (2020) find that some contributors often invest in more than one campaign, and such serial investors contribute earlier. However, they are not more informed and fail to pick better quality ICOs. On the other hand, Hornuf, Kück and Schwienbacher (2021) show that issuers who disclose their source code are more likely to be targeted by hackers and scammers, highlighting the risk of disclosing the code. They find it extremely difficult to predict fraud with the information available (whitepapers and other sources like websites, social media accounts, etc.) at the time of ICO issuance. Zhang, Aerts, Lu, and Pan (2019) study the data from the four largest tokens exchanges in Asia and their findings indicate that whitepapers with more readable disclosures are likely to result in a higher first-day return.

2.2. Investors' sentiments, whitepaper readability, and ICO performance

Baker and Wurgler (2007) use several market-based measures as proxies for investor sentiment. Besides the market-based measure, the other most common approach applied in earlier research has been survey-based indices. More recently, building an investor-sentiment index employing qualitative transcripts (for example 10K filings, whitepapers, earning announcements, etc), daily news, internet search, social media content, blogs, etc. have gained popularity because traditional approaches like market-based and survey-based methods seem to be less transparent.

Drobetz, Momtaz, and Schröder (2019) examine to what extent the market for ICOs is driven by investor sentiment. Their results, based on sentiment and coin price data, show that the ICO market is driven by digital financial market sentiment, whereas it is almost unrelated to general capital market sentiment. Their results show that social media channels overdrive traditional news channels as the main source of investor sentiment. Similarly, Domingo, Piñeiro-Chousa, and López-Cabarcos (2020) also find that sentiment extracted from social networks positively influences ICO returns. Specifically, the authors document that Bitcoin spot and Bitcoin futures returns are positively correlated with ICO returns, whereas the existence of a presale period has a negative influence. Zhang, Aerts, Lu, and Pan (2019) study the linkage between the readability of whitepapers and the first-day return. Using data from the four largest tokens exchanges in Asia, their findings indicate that whitepapers with more readable disclosures are likely to result in a higher first-day return for ICO investors. Similarly, Qadan (2019) uses readily available 11 different sentiment indices as different proxies of risk appetite. These indices are Baker and Wurgler's (2006) index; Huang et al. (2015) HJTZ Index; Baker et al. (2016) Economic /Monetary Policy Uncertainty Index (EPU); American Association of Individual Investors' (AAII) Sentiment Survey, Consumer Sentiment Index (CSI); Consumer Confidence Index(CCI); Louis Fed Financial Stress Index (STLFSI). This is not the same in our case as we quantified the qualitative data (whitepapers) using natural language processing tools and extracted sentiment scores by applying four different sentiment dictionaries and further expand it to the emotional level. We also extracted the seven different readability scores from each of the ICO whitepapers.

2.3. Connectivity of CEO and the advisors and ICO performance

Giudici, Moncayo, and Martinazzi (2020) find a relationship between the number of advisors' connections and their capability. They conclude that advisors in connection with multiple ICOs bridge the gap between the network and result in ICO success. They also show that the well-connected advisors in other ICOs are directly related to a larger amount of raised funding. Similarly, Amsden and Schweizer (2018) study 1,009 ICOs from 2015 to March 2018 and highlight those better-connected CEOs are positively correlated with ICO success. Moreover, providing information on a hard cap in a pre-ICO can help investors measure success in the pre-sale. Momtaz (2020) explores the factors affecting the ICO market outcomes and finds that management quality and project quality are positively correlated with the funding amount and returns. This study also finds that highly visionary projects harm success. Furthermore, the study shows that highly visionary projects are more likely to fail, as 21% of all tokens got delisted from a major exchange platform during the sample period.

[6] In November 2017, the European Supervisory Markets Authority issued statements notifying investors and firms of potential risks native to certain ICOs. The authority notified that certain feature ICOs may be governed by existing EU legislation.

2.4. Technical factors among ICOs and their signaling capabilities

Analyzing a data of 1,392 projects, Yu (2019) shows that the volatility of the main cryptocurrencies has a significant impact on the success of ICOs. For example, the success of ICOs on smart contracts built upon the ERC-20 token primarily depends on the volatility of Ethereum and secondarily on all other factors such as team quality. Furthermore, Meyer and Ante (2020) analyze 250 cross-listings of 135 different tokens and calculate abnormal returns for specific samples using an event study. They find that returns are driven by success in terms of token performance and project funding as well as characteristics such as regulation and domestic market size of the ICO issuing party.

Other characteristics such as blockchain infrastructure, token distribution, team, campaign duration, and whitepaper characteristics also seem to influence the perceived project quality as well as the cross-listing returns. Fisch (2019) assesses the determinants of the amount raised in 423 ICOs. The study explores the role of technological capabilities among ICOs and their signaling capabilities. The results also show that technical whitepapers and high-quality source codes are positively related to the amount of raised funding. Surprisingly, patents and copy rights, which can be considered quality signals, are not associated with increased amounts of funding.

3. Data And Methodology

3.1. Preparing the data set

We applied *rvest* and *xml2* web scrapping packages in the standard statistical software R to download the data from icorating.com., icosbull.com, and tokendata.io. Icorating.com has the *risk score* (e.g., a score for measuring potential fraud) and the *hype score* (e.g., a score resembling the number of social media followers) for more than 5,000 ICOs with additional information on the amount of raised funding which we denote in our study as *raised* measured in terms of USD. Similarly, the website icosbull.com provides basic data (i.e., data such as the *Name, Symbol, Description, Country*), financial data (i.e., data such as *Softcap, Hardcap, Raised Amount*) and data on social signal views (i.e., data such as Telegram or Twitter followers covering around 3,000 ICOs.) Moreover, the website tokendata.io has information on the daily price and return data on the token sales of the ICOs including the raised amount. We downloaded the financial information of those listed ICOs using the same web scrapping packages. Unfortunately, the financial information for many ICOs is missing on these three websites. Furthermore, financial information, especially on the raised amount of funding, which is of major importance in our study, is missing on many websites (including major ICO database providers such as icobench, neironix, icoholders) Nevertheless, after combining data retrieved from icorating.com, iscosbull.com and tokendata.io, we were able to collect the information about the raised amount for 1,507 ICOs issued in the August 2014 to December 2019 period. Furthermore, we also observe some non-uniformity in the reported raised amount for some completed ICOs on these websites. Most of the reported raised amounts on these websites are rounded in thousands and millions. If available, our sample takes the exact raised amount (not rounded) from the above-mentioned websites.

As a result, we manually collected all 5,033 ICO whitepapers from various sources. Neironix.io provides access to the majority of ICO whitepapers by providing a direct link to the website of the ICO providers. However, some failed, scam and unsuccessful ICOs unfortunately removed whitepaper access from their websites. Fortunately, some websites have copies of whitepapers in their own databases. Using intensive manual work, we are able to collect the whitepapers of all unique ICOs for this study. Figure 1 shows the step-by-step process of the data retrieval highlighting various sources used to obtain the data set.

Figure 2 and Figure 3 show the geographical representation of amount raised by ICOs in US dollar and the number of ICOs launched during 2014-2019 period respectively. Furthermore, Figure 4 shows the pie chart of ICOs registered under top 20 different categories.

From data retrieval to tabulation of the results, we used various R packages which, along with their functions and usage, are described in **Appendix A.1.**

3.2. What variables could we identify after combining the information from whitepapers with other data sources?

3.2.1. How long, detailed and accessible is the whitepaper?

In total, we were able to identify 36 different variables for 1,507 ICOs of which we have information on the raised amount of funding. For each ICO published and completed during 2014-2019 we gathered project related characteristics mainly from the particular ICO's whitepaper and also from different open source websites listing and rating ICOs (see Figure 1). We report the definitions for our variables in Table 1 and the corresponding summary statistics of those variables in Table 2.

We list *Word Count* as a possible predictor variable affecting the raised amount as a detailed and inclusive whitepaper might have an extensive word count. It is important to understand whether or not a detailed whitepaper is sufficient to raise the amount needed to fund the business project of the ICO. Our data set reveals that there is one whitepaper with more than 41 thousand words and at the same time we observe that the minimum of words used in a whitepaper is 241 words only. On average, a whitepaper in our sample includes 4,338 words. In this regard, Figure 5 displays a word cloud for the most frequently used words in these whitepapers. It shows that many ICO whitepapers are focused on token, usage, blockchain, and platform.

Similarly, the *Page Length* of an ICO may also provide some useful information on the project. Besides words, a page also consists of many graphical explanations. A lengthier whitepaper tends to have a large number of words and may also include many tables and graphs. Some ICO projects provide a light paper i.e., *'one-page white paper'* instead of or even on top of a regular whitepaper. There are many ICOs in our data sample that provide only the lightpaper version of their whitepaper. The lengthiest whitepaper in our data sample exhibits almost 300 pages, whereas the shortest one is just a single page lightpaper. Mostly, the whitepapers are A4 and Letter

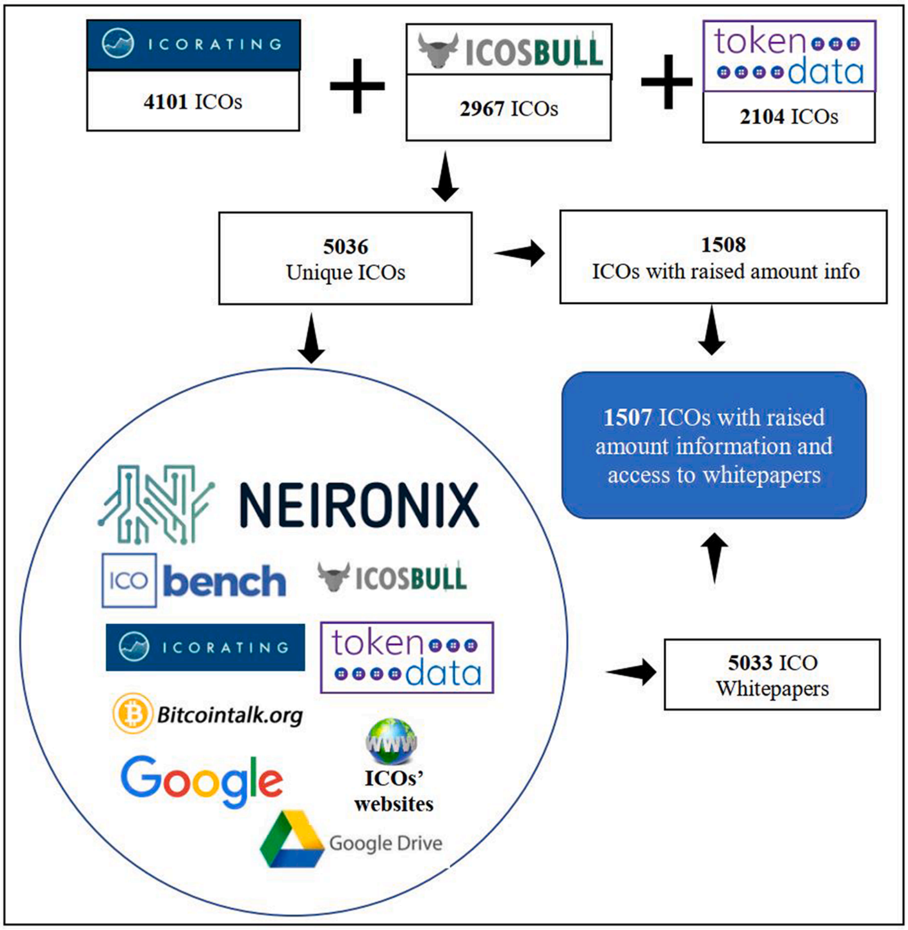


Fig. 1. ICO data accumulation process (2014-2019). This figure shows our sample of 1508 ICO with raised amount info data generation process.

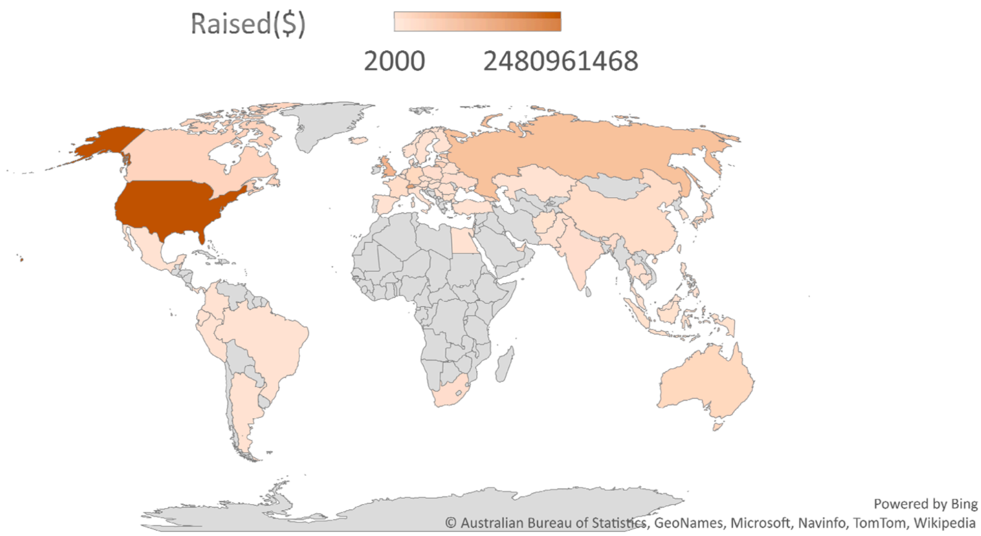


Fig. 2. Geographical representation of amount raised by ICOs during 2014-2019 (This heat map is created using Microsoft Excel, it includes 1507 ICOs with Raised amount available, for actual USD figures see **Appendix A.2**. and for the evolution of the funds raised using ICOs over time see **Appendix A.3.**) *Note: This map excludes raised amount data for 765 ICOs where the country information is not disclosed.*

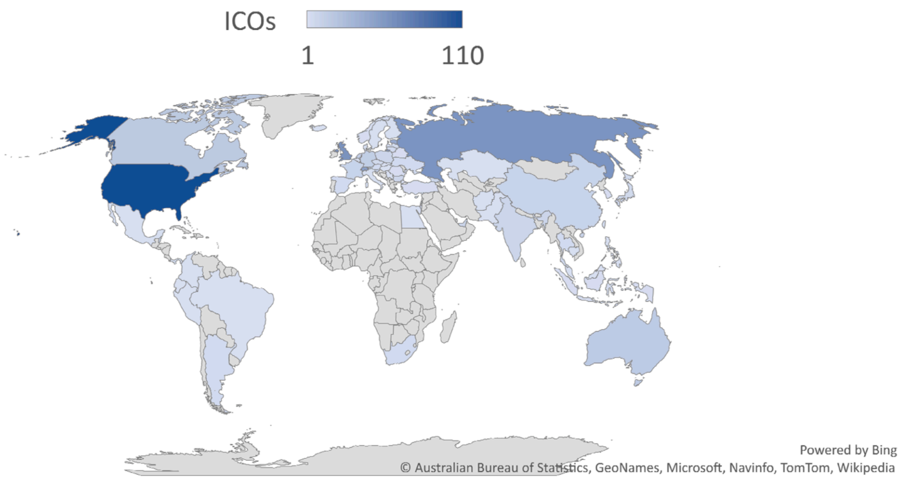


Fig. 3. Geographical representation of number of ICOs launched during 2014-2019 (This heat map is created using Microsoft Excel, it includes 1507 ICOs with Raised amount available, for exact numbers of ICOs, see **Appendix A.2.**) *Note: This map excludes 765 ICOs where the country information is not disclosed.*

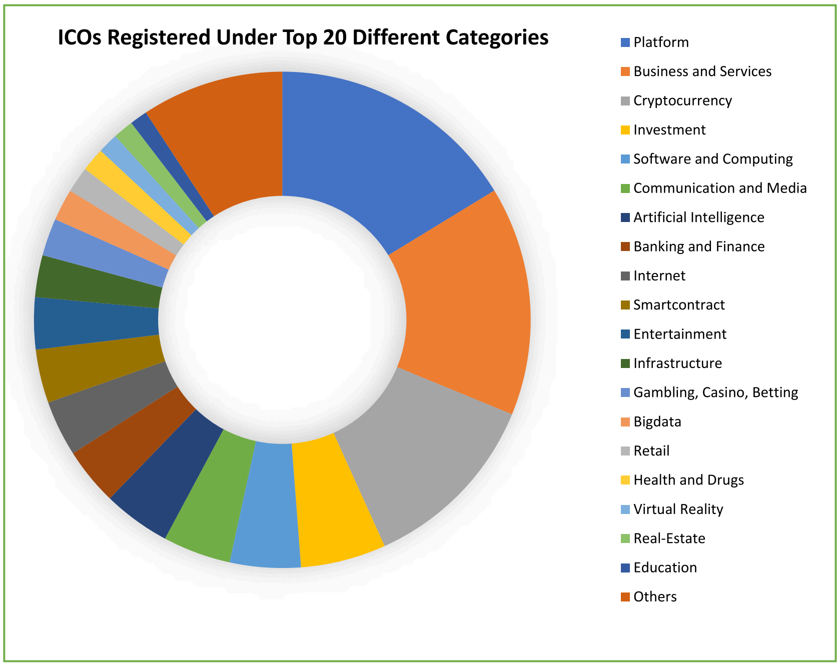


Fig. 4. Pie chart representing ICOs registered under top 20 categories (Note: same ICO is registered under different categories, for detail see **Appendix A.4.**)

size files in portable document format (pdf). However, some whitepapers are a single long pdf file with no page division. Our methodology treats these files as single page files by default.

Readability score measures the reading difficulty of a document. The question arises whether it is difficult (easy) to raise funding if the whitepaper is difficult (easy) to read. We seek to answer this question by applying different readability measures in our studies[7]. For example, the Flesch Readability score for 1,507 whitepapers exhibiting information on raised funding has a mean of 34.461, which shows that, on average, whitepapers are difficult to read and best understood by college graduates (see **Appendix A.7.**). There is no limit on the lowest side of the Flesh readability score. Very complicated sentences can have negative scores. The lowest Flesh score in

[7] To understand how the readability scores for different readability measures are calculated, see **Appendix A.6.**

Table 1
Variable Names, Descriptions and Their Sources

S. No.	Variables	Full Name	Description	Sources
1	Name	Name of the ICO	ICOs published since 2014-2019	icorating.com, icobulls.com, tokendata.io
2	PL	Page Length	total number of pages in a whitepaper(pdf) file	icorating.com, icobench.com, icobulls.com, neironix.io
3	WC	Word Count	total number of words in a whitepaper corpus	Corpus, R SentimentAnalysis Package
4	S.GI	Overall Sentiment GI	psychological Harvard-IV dictionary (Overall= positive-negative)	Corpus, R SentimentAnalysis Package
5	N.GI	Negative Sentiment GI	psychological Harvard-IV dictionary (Negative)	Corpus, R SentimentAnalysis Package
6	P.GI	Positive Sentiment GI	psychological Harvard-IV dictionary (Positive)	Corpus, R SentimentAnalysis Package
7	S.HE	Overall Sentiment HE	Henry's Business Communication dictionary (Overall= positive-negative)	Corpus, R SentimentAnalysis Package
8	N.HE	Negative Sentiment HE	Henry's Business Communication dictionary (Negative)	Corpus, R SentimentAnalysis Package
9	P.HE	Positive Sentiment HE	Henry's Business Communication dictionary (Positive)	Corpus, R SentimentAnalysis Package
10	S.LM	Overall Sentiment LM	LoughranMcDonald finance-specific dictionary (Overall=positive-negative)	Corpus, R SentimentAnalysis Package
11	N.LM	Negative Sentiment LM	LoughranMcDonald finance-specific dictionary (Negative)	Corpus, R SentimentAnalysis Package
12	P.LM	Positive Sentiment LM	LoughranMcDonald finance-specific dictionary (Positive)	Corpus, R SentimentAnalysis Package
13	RU.LM	Uncertain Sentiment LM	Sentiment uncertainty LoughranMcDonald finance-specific dictionary (neither Negative nor Positive)	Corpus, R SentimentAnalysis Package
14	S.QDAP	Overall Sentiment QDAP	QDAP Qualitative Data Analysis Program University of Pittsburgh (Overall=positive-negative)	Corpus, R SentimentAnalysis Package
15	N.QDAP	Negative Sentiment QDAP	QDAP sentiment polarity of text by grouping variables (Negative)	Corpus, R SentimentAnalysis Package
16	P.QDAP	Positive Sentiment QDAP	QDAP sentiment polarity of text by grouping variables (Positive)	Corpus, R SentimentAnalysis Package
17	S.Score	SimilarityScore	Jaccard Similarity Socer (25,331,089 pair comparision for similarity) 2nd highest match	Corpus, R Textreuse Package
18	Flesch	ReadabilityScoreType1	Readability score selected among the several measures where R output displays no errors	Corpus, Readability, KoRpus Package
19	Flesch.K	ReadabilityScoreType2		
20	RIX	ReadabilityScoreType3		Corpus, Readability, KoRpus Package
21	SMOG	ReadabilityScoreType4		Corpus, Readability, KoRpus Package
22	FOG	ReadabilityScoreType5		Corpus, Readability, KoRpus Package
23	ARI	ReadabilityScoreType6		Corpus, Readability, KoRpus Package
24	Col	ReadabilityScoreType7		Corpus, Readability, KoRpus Package
25	HScore	HypeScore	Hype Score (Dummy, Rated(1) not rated (0))	icorating.com
26	RScore	RiskScore	Risk Score (Dummy, Rated(1) not rated (0))	icorating.com
27	NoC	NumberofCategory	ICO registered under different categories(1-14)	Icobulls.com
28	Team	TeamInformation	Team size	icorating.com
29	C.Info	CountryInformation	Country of issue information Dummy (1 given, 0 not given)	Icobulls.com
30	KYC	KnowYourCustomer	Known Your Customer Scor	icorating.com
31	Twitter	Twitter Information	Social media profile followers	icorating.com
32	Miles	Milestones	Milestone information in website/corpus Dummy (1 given, 0 not given)	Corpus, icorating.com
33	SoftCap	SoftCapital	Softcap information website/database/corpus Dummy (1 found, 0 not found)	Corpus, icorating.com
34	HardCap	HardCapital	Hardcap information website/database/corpus Dummy (1 found, 0 not found)	Corpus, icorating.com
35	NoT	Number of Tokens Info	Number of tokens Dummy (1 known, 0 unknown)	icorating.com
36	RaisedA	Raised Amount Info	Raised amount information in USD	icorating.com, icosbull.com, tokendata.io

our sample is -75.58. Flesh-Kincaid Grade Level is a readability test designed for English texts. Note that the test focuses on polysyllabic words and long sentences. It measures reading difficulty related to the approximate US grade level.

Similarly, Automated Readability Index (ARI) works well with both English and Western European languages. It uses long words and long sentences to calculate a readability score. It indicates how difficult the page is to read. The score can be matched to an equivalent reading ability level. The mean ARI of 14 in our sample in Table 2. implies that on average the whitepaper can be read and understood by 14th grade (i.e., university degree) students. Furthermore, Coleman-Liau is another recognized readability test designed primarily for English texts. It focuses on long words and long sentences. Using this test, a score can only be calculated if the content exceeds 100 words. Since the lowest number of words in our data sample of whitepapers is 288, we have no difficulties in using this readability measure. The test produces an approximate representation of the US grade level needed to comprehend the text.

Gunning Fog is a readability test for English texts only. It also focuses on complex polysyllabic words and long sentences. We also include the SMOG readability test which is developed just for English texts. This measure also primarily focuses on polysyllabic words. However, a score can only be calculated if the document has at least 30 sentences. Another readability measure is the Rate Index RIX created by the Australian teacher Jonathan Anderson. Anderson wanted to convert the formula to a grade level. The average score of 7.5 (7.2 and above = 12^{th} grade) in RIX tells us that the average whitepaper can be read by college level students. Unlike other

Table 2
Summary Statistics

Variables	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(50)	Pctl(75)	Max
Word Count	1507	4337.591	2835.686	241	2411.5	3820	5645	41532
Page Length	1507	31.999	17.936	1	20	29	41	288
Flesch Readability Score	1507	34.461	11.503	-75.576	29.137	34.95	40.883	120.205
Flesch.K Readability Score	1507	14.629	3.462	-3.010	12.828	14.25	15.825	66.863
RIX Readability Score	1507	7.457	2.586	0.000	6.056	7.161	8.362	45.136
SMOG Readability Score	1507	15.613	1.936	3.129	14.430	15.467	16.606	32.506
FOG Readability Score	1507	17.861	3.643	0.800	15.957	17.49	19.150	71.530
ARI Readavility Score	1507	15.160	4.461	-6.300	12.934	14.71	16.570	87.841
Col Readability Score	1507	34.007	5.781	-38.450	30.775	33.44	36.692	90.550
Sentiment GI	1507	0.142	0.038	-0.013	0.124	0.143	0.166	0.240
Negative GI	1507	0.064	0.018	0.000	0.054	0.063	0.073	0.148
Positive GI	1507	0.206	0.040	0.0002	0.191	0.209	0.227	0.300
Sentiment HE	1507	0.013	0.008	-0.017	0.008	0.013	0.017	0.096
Negative HE	1507	0.006	0.004	0.000	0.003	0.005	0.008	0.030
Positive HE	1507	0.019	0.007	0.000	0.014	0.018	0.023	0.099
Sentiment LM	1507	-0.014	0.016	-0.088	-0.024	-0.014	-0.004	0.050
Negative LM	1507	0.045	0.015	0.000	0.036	0.045	0.055	0.107
Positive LM	1507	0.031	0.010	0.000	0.026	0.031	0.037	0.090
RUncertainty LM	1507	0.013	0.006	0.000	0.008	0.012	0.016	0.061
Sentiment QDAP	1507	0.100	0.029	-0.027	0.084	0.102	0.119	0.197
Negative QDAP	1507	0.036	0.013	0.000	0.028	0.035	0.044	0.109
Positive QDAP	1507	0.137	0.029	0.000	0.125	0.139	0.153	0.225
Raised Amount (1M, USD)	1507	20.265	122.517	0.0001	2.314	8.549	19.956	4234.276
Jaccard Similarity Score	1507	0.061	0.130	0.000	0.010	0.02	0.050	1.000
KYC Score	1507	1.656	1.767	0	0	1,655	3.4	5
Twitter Followers	1507	3897.20	16428.26	0	0	0	2103.20	318271
No. of Categories	1507	1.652	1.307	1	1	1	2	11
Team Size	1507	4.819	7.490	0	0	4	9	45
Dummy Variables	N	Mean	St.Dev.					
High Hype Score	1507	0.151	0.358					
Medium Hype Score	1507	0.280	0.449					
Low Hype Score	1507	0.248	0.432					
Hype Score Not Rated	1507	0.292	0.455					
High Risk Score	1507	0.111	0.315					
Medium Risk Score	1507	0.182	0.386					
Low Risk Score	1507	0.062	0.241					
Risk Score Not Rated	1507	0.639	0.481					
Country Disclosed	1507	0.493	0.500					
Road Map/ Milestone	1507	0.716	0.451					
SoftCap Disclose	1507	0.236	0.425					
HardCap Disclose	1507	0.309	0.462					
Number of Tokens Disclose	1507	0.310	0.463					

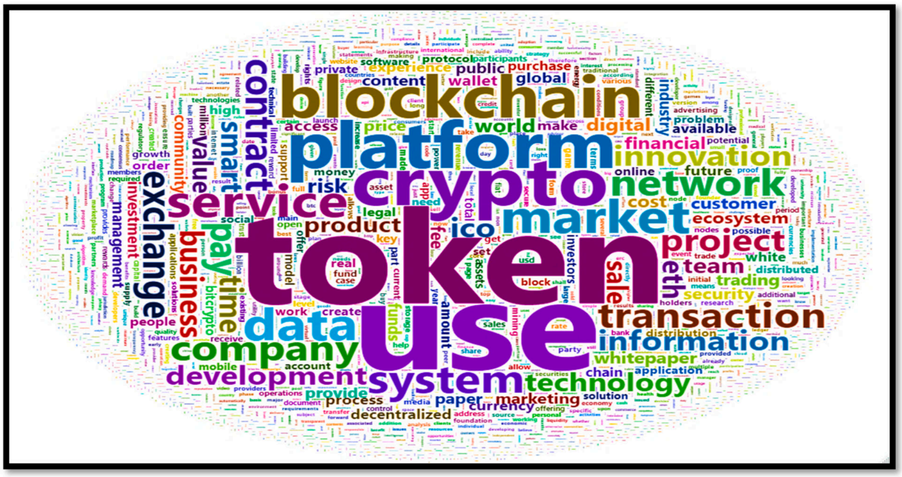


Fig. 5. Word cloud of 5033 ICO whitepapers 2014-2019 (This word cloud is created using the wordcloud2 package in R). (**Note:** To see the frequency of the words, please see **Appendix A.5.**)

readability measures, RIX can be used for both English and non-English texts.

3.2.2. Identifying the sentiment in whitepapers using popular sentiment dictionaries

Another important novel element in our study is the sentiment hidden in the text of an ICO whitepaper. By applying four different sentiment dictionaries we seek to answer if positive/negative sentiment is associated with success/failure in raising capital for ICOs. We apply the *SentimentAnalysis* package in R statistical software, giving us positive, negative and overall sentiment scores for four different sentiment dictionaries applied to ICO whitepapers. In this regard, the *AnalyzeSentiment* function from the above-mentioned package gives sentiment scores for:

i. *Harvard-IV General Purpose Psychological Dictionary*
ii. *Henry's Finance Dictionary*
iii. *Loughran Mc Donald Finance-specific Dictionary*
iv. *QDAP (Quantitative Discourse Analysis Package) Dictionary*

Firstly, the Harvard-IV psychological dictionary, comprising a list of positive and negative words, is a general-purpose dictionary developed by the Harvard University. It maps each whitepaper with counts on positive, negative and overall sentiment[8]. The mean result in the summary data statistics shows that the whitepapers have slightly more positive psychological sentiment then negative.

Second, Henry (2008) studies the genre of earnings press releases along with quantitative analysis. He uses capital markets data to assess the investor impact of tone and other stylistic attributes. The study's results suggest that tone influences investors' reactions. Using the dictionary developed by the author, the R package gives us the sentiment scores as positive, negative and overall. Furthermore, we get the similar sentiment variation with this dictionary too, where the mean positive sentiment score is slightly larger than the mean of negative sentiment.

Third, Loughran and McDonald (2011) shows that word lists developed for other disciplines misclassify common words in financial text. Employing a large sample of 10-Ks during the 1994 to 2008 period, they find that almost three-fourths of the words identified as negative by the widely used Harvard Dictionary are words typically not considered negative in financial contexts. This is also noticeable in our summary statistics outlined in Table 2**.** Specifically, applying this dictionary, the negative sentiment score is higher than the positive sentiments for the ICO whitepapers, which gives us the negative overall sentiment. Furthermore, this measure also shows the equal positive risk for sentiment uncertainty[9]. It is interesting to see whether the dictionaries developed for the finance specific areas can capture the sentiment of this new digital financial market.

Fourth, the Quantitative Discourse Analysis Package (QDAP) provides quantitative analysis of qualitative transcripts and therefore bridges the gap between quantitative and qualitative research approaches. It overlaps with natural language processing and text mining. We also get the similar pattern using QDAP, that is, on average, positive sentiment is higher than negative sentiments in the sample of ICO whitepapers.

3.2.3. On encountering Déjàvu: Is the whitepaper subject to plagiarism?

Next, we are interested in analyzing whether providing a similar whitepaper in terms of tokenized contents (words/sentences/paragraphs) is either beneficial or harmful for the projects in terms of attaining funding. Similarities in whitepapers can be beneficial to the investors as familiar words and sentences may increase confidence. On the other side, investors can accuse the project of plagiarism, if the content is highly similar. In our study, we use the *jaccard similarity* (TextReuse R package) as a proxy for the similarity of whitepapers. The function jaccard_similarity provides the Jaccard measures of similarity for two sets. The coefficients will be numbers between 0 and 1. For the similarity coefficient, the higher the figure the more similar the two sets of whitepapers. We make over 25 million pairwise comparisons (5,033x5,033= 25,331,089) and get bi-directional comparisons. Note that we did not use the function *jaccard_bag_similarity* which provides only one directional comparison. We found that, on average, one ICO whitepaper is around 6% similar to all the other whitepapers. After excluding the self-matched pair there are still some ICO whitepapers fully copied from another project with a different name, which have a Zaccard similarity score of 1.

3.2.4. Media attention and social media

Another important variable that we consider is the *Hype Score,* which shows the level of interest in the project from potential investors. The values High, Medium and Low is calculated based on the number of users on the project pages on social media. Social media includes the websites *bitcoin forum* and *telegram*. The higher the value, the more people are interested in the project, which indicates a potentially high demand for the tokens. This score is available in icorating.com for free for the majority of planned, ongoing and completed ICOs. We use dummy variables to indicate the level of *Hype Score* for high, medium and low as well as for those ICOs with no rated Hype Score. The summary statistics reported in Table 2**.** shows that 15.1% of ICOs have high, 28% have medium, 25% have low and 32.1% of ICOs have no rated hype score. This implies that the minority of the ICOs have a large number of social media followers.

Nowadays, social media like Telegram and Bitcoin forum are important tools for communication and Twitter is valuable for mass communication. The number of twitter followers of a respective ICO project can indicate the popularity of the project. We consider

[8] For more please visit; http://www.wjh.harvard.edu/~inquirer/

[9] Overall sentiment score for Loughran and McDonald = -0.014 and sentiment uncertainty risk = 0.013 (see Table 2.)

number of *Twitter Followers* as a potential variable that can be associated with the raised amount of funding. We mark ICOs with no Twitter account and assign them 0 followers. The summary statistics reported in Table 2**.** show that, on average, one ICO has around 4,000 Twitter followers.

3.2.5. Risk, disclosure, industrial sectors

Similar to the *Hyper Score,* the website icorating.com also provides the *Risk Score* of the project. It is used to assess risks of potential fraud, as well as the overall quality of project development. This variable determines the reliability of a project against aspects such as its team, the product, the existence of partners and so on. The value from low to high shows the risk of fraud from small to large. This is also a measure for the project's investment attractiveness. This score is available in icorating.com for free for the majority of planned, ongoing and completed ICOs. Again, we use dummy variables to indicate the level of risk score for high, medium, low and also for those ICOs whose *Risk Score* is not rated. Our summary statistics reported in Table 2**.** show that 11% of ICOs have high, 18% have medium, 7% have low and 64% of ICOs have no rated risk score. Even ICOs exhibiting no rated risk score can have valuable information because these ICOs have relatively low level of information disclosure.

Next, we extracted a Know Your Customer (KYC) score from all the ICOs with information on the raised amount from the popular ICO website ICObench.com. KYC is an essential method of verifying the identity of the teams and people. ICObench employs this procedure to verify the identity of the ICO/IEO team members and to verify experts and bloggers. Many bounty hunters participate in KYC campaigns in exchange for free tokens. However, a high KYC score is also a symbol of higher level of trust in the project. The highest score of 5 shows that the project is highly reliable whereas a KYC campaign not undertaken (score=0) shows that there is a high risk that the project has fake teams and other stakeholders. The summary statistics reported in Table 2**.** indicate that our data sample exhibits a mean KYC score of 1.66, which is relatively low.

Next, in our list of variables, the *Number of Categories* is also a potentially important predictor variable. Investors might be interested in investing in a particular sector due to having prior knowledge in that specific industry. For instance, considering utility tokens, investors prefer to use platforms or services in some particular market segment. In this regard, it is important to note that one single ICO can be registered under multiple categories. As an example, our sample exhibits one ICO which has been registered in 11 different categories. Figure 4 shows the top 20 different categories wherein ICOs have been registered. It is important to explore whether identifying an ICO as ICO related to some particular sector or general purpose ICO provide information on the success in raising funding. Hypothetically, innovative business sectors related to platforms could be more appealing to the investors than finance sectors which can be considered traditional business sectors.

3.2.6. Where can we find you? The location of the project team

Another important variable that we consider is the team size. Some ICOs have a very large team size. For instance, the maximum number of people in a team in an ICO in our data sample is 45 (see Table 2**.**). The question arises whether a larger team is of support for increasing the amount of raised crowdfunding. Hypothetically, a larger group of developers could be considered more competent than a smaller group of developers. We also seek to answer this question by considering this variable in our analysis. ICOs with no team information disclosure are assigned 0 team members.

Further, we consider some important dummy variables in our regression model. Some ICO issuers do not disclose the issuers' home country. This information might be crucial to the investors as ICOs from certain countries have been revealed as scam due to poor domestic financial regulation. The Transparency Index of a country measuring the level of corruption and the gross domestic product as a measure for the level of economic activities might be an indication of the trust and future prospect of an ICO. Only 50% of ICOs in our data sample have disclosed their home country. Hypothetically, concealing the country of origin could indicate a deliberate act for possible fraud. If the ICO has disclosed the country of issuance it receives a dummy variable of 1 and otherwise a variable of 0. The website icosbull.com provides information on the issuing country of past ICOs.

3.2.7. How does the ICO want to proceed and what does this progress require?

Next, we assign a dummy variable for those ICOs that have clear *Roadmap/Milestones* mentioned in their whitepaper. A roadmap or milestone shows the future prospect of the ICO and, hence, may serve as strong quality signal. A clear future plan might be the indication of a legitimate ICO. 72% of ICOs in our data sample exhibit a clear roadmap/milestone that is elaborately mentioned in the whitepaper (see Table 2**.**).

Moreover, we define the dummy variables *softcap* and *hardcap.* Specifically, the softcap is the lower limit and the hardcap is the upper limit of the required funding for an ICO. If a team receives funding exceeding the hardcap, it should be returned to the investors. Failing to do so is a red flag for possible scam. Disclosure of required capital helps investors to monitor the progress of the ICO in collecting funds. Many ICO issuers do not disclose this information, meaning that investors do not have any prenotion of whether the project is going to be successful. Hiding this essential information could hypothetically help to generate a continued flow of funds as the investors are unaware of whether the required capital limit is reached. Only around one fourth of the ICOs have disclosed softcap information whereas around one third have disclosed the hardcap information in our data sample. Again, it might be an indication of scam if they are withholding the information of soft and hard capital. The website icosbull.com provides data on the softcap and hardcap (see the information on the financial view of the past ICOs). Similarly, icorating.com has the information on the capital goal. We also used text mining to see if the softcap and hardcap information exists in the whitepaper. ICOs with this information disclosure are allocated a dummy variable of 1 and those without receive a dummy variable of 0.

Finally, the *Number of Tokens Disclosed* is the last important dummy variable used in our analysis. This variable exhibits crucial information, that is, either the token is heavily distributed among developers during the pre-sale phase or not. One can find the predistribution percentage by simply taking the ratio of pre-sale tokens and total number of tokens. If the total number of tokens is not disclosed beforehand, it could be an indication of scam (i.e., 'pump-and-dump'). In this regard, Grobys and Sapkota (2020) study potential determining factors for cryptocurrency default and they show that high levels of pre-mining could potentially be a get rich quick scheme on the part of the developers, rather than setting up the coin for long-term success and ultimately leading to default. Descriptive statistics (Table 2**.**) show that only one third of the ICOs have disclosed information on the number of tokens.

3.3. Statistical model

In this paper we focus on three major features associated with ICOs. The first is the level of information disclosures in terms of availability of necessary information on the whitepaper itself. We address this feature by using variables such as Roadmap/Milestone, softcap, hardcap and disclosure of tokens numbers.

Second, this paper quantifies the qualitative aspects of whitepapers such as sentiment and readability. We address this feature by using four different sentiment measures following *(i) Harvard-IV General Purpose Psychological Dictionary, (ii) Henry's Finance Dictionary, (iii) Loughran Mc Donald Finance-specific Dictionary* and, *(iv) QDAP* in association with seven different readability indices, which are, *(i) Flesh, (ii) Flesh-Kincaid., (iii) ARI, (iv) Coleman-Liau, (v) Gunning Fog, (vi) SMOG*, and *(vii) RIX*.

Third, we also account for the characteristics of the ICO project found outside of the whitepapers such as, social media followers, possible scams and KYC score. We apply the multiple linear regression model based on pooled ordinary least squares for the parameter estimation given by equations (1) to (4) for four sentiment dictionaries.

$$\begin{aligned}\ln\left(Raised_{i,j}\right) = {} & \beta_0 + \beta_1 WC_i + \beta_2 PL_i + \beta_3 N.GI_i + \beta_4 P.GI_i + \beta_5 Read_{i,j} + \beta_6 Sim_i + \beta_7 KYC_i + \beta_8 Twt_i + \beta_9 Team_i + \beta_{10} H.H_i + \beta_{11} M.H_i \\ & + \beta_{12} L.H_i + \beta_{13} H.NR_i + \beta_{14} H.R_i + \beta_{15} M.R_i + \beta_{16} L.R_i + \beta_{17} R.NR_i + \beta_{18} CD_i + \beta_{19} NoC_i + \beta_{20} RM.MS_i + \beta_{21} SC_i + \beta_{22} HC_i \\ & + \beta_{23} NoT_i + \in_i \end{aligned} \tag{1}$$

Equation (1) is the regression model employing the Harvard-IV General Purpose Psychological Dictionary. Moreover, $\ln\left(Raised_{i,j}\right)$ is the log of raised amount for each ICO i, for seven different readability measures j. The independent variables in this equation are WC_i (word count), PL_i (page length), $N.GI_i$ (negative Harvard psychological sentiment), $P.GI_i$ (positive Harvard psychological sentiment), $Read_{i,j}$ (seven different Readability measures from model (1) - model (7)), Sim_i (similarity score), KYC_i (know your customer score), Twt_i (Twitter followers), $Team_i$ (team size), $H.H_i$ (hype score dummy, high), $M.H_i$ (hyper score dummy, medium), $L.H_i$ (hype score dummy, low), $H.NR_i$ (hype score dummy, not rated), $H.R_i$ (risk score dummy, high), $M.R_i$ (risk score dummy, medium), $L.R_i$ (risk score dummy, low), $R.NR_i$ (risk score dummy, not rated), CD_i (country disclosure dummy), NoC_i (number of categories), $RM.MS_i$ (roadmap/milestone disclosure dummy), SC_i (softcap disclosure dummy), HC_i (hardcap disclosure dummy), NoT_i (number of tokens disclosure dummy).

In equation (2), our regression model accounts for Henry's Finance Dictionary, where, $N.HE_i$ is scores for Henry's negative finance sentiment and $P.HE_i$ is scores for Henry's positive finance sentiment.

$$\begin{aligned}\ln\left(Raised_{i,j}\right) = {} & \beta_0 + \beta_1 WC_i + \beta_2 PL_i + \beta_3 N.HE_i + \beta_4 P.HE_i + \beta_5 Read_{i,j} + \beta_6 Sim_i + \beta_7 KYC_i + \beta_8 Twt_i + \beta_9 Team_i + \beta_{10} H.H_i + \beta_{11} M.H_i \\ & + \beta_{12} L.H_i + \beta_{13} H.NR_i + \beta_{14} H.R_i + \beta_{15} M.R_i + \beta_{16} L.R_i + \beta_{17} R.NR_i + \beta_{18} CD_i + \beta_{19} NoC_i + \beta_{20} RM.MS_i + \beta_{21} SC_i + \beta_{22} HC_i \\ & + \beta_{23} NoT_i + \in_i \end{aligned} \tag{2}$$

Similarly, in equation (3), we employ the Loughran Mc Donald Finance-specific Dictionary to assess the whitepaper sentiment measures,

$$\begin{aligned}\ln\left(Raised_{i,j}\right) = {} & \beta_0 + \beta_1 WC_i + \beta_2 PL_i + \beta_3 N.LM_i + \beta_4 P.LM_i + \beta_5 RU.LM_i + \beta_6 Read_{i,j} + \beta_7 Sim_i + \beta_8 KYC_i + \beta_9 Twt_i + \beta_{10} Team_i + \beta_{11} H.H_i \\ & + \beta_{12} M.H_i + \beta_{13} L.H_i + \beta_{14} H.NR_i + \beta_{15} H.R_i + \beta_{16} M.R_i + \beta_{17} L.R_i + \beta_{18} R.NR_i + \beta_{19} CD_i + \beta_{20} NoC_i + \beta_{21} RM.MS_i + \beta_{22} SC_i \\ & + \beta_{23} HC_i + \beta_{24} NoT_i + \in_i \end{aligned} \tag{3}$$

where, $N.LM_i$ defines the negative sentiment as measured via Loughran Mc Donald sentiment dictionary, $P.LM_i$ defines the corresponding positive sentiment, and $RU.LM_i$ defines the measure for the corresponding sentiment uncertainty risk.

Furthermore, we also account for the Qualitative Discourse Analysis Package (QDAP) for sentiment polarity in the equation (4) below.

$$\begin{aligned}\ln\left(Raised_{i,j}\right) = {} & \beta_0 + \beta_1 WC_i + \beta_2 PL_i + \beta_3 N.QDAP_i + \beta_4 P.QDAP_i + \beta_5 Read_{i,j} + \beta_6 Sim_i + \beta_7 KYC_i + \beta_8 Twt_i + \beta_9 Team_i + \beta_{10} H.H_i \\ & + \beta_{11} M.H_i + \beta_{12} L.H_i + \beta_{13} H.NR_i + \beta_{14} H.R_i + \beta_{15} M.R_i + \beta_{16} L.R_i + \beta_{17} R.NR_i + \beta_{18} CD_i + \beta_{19} NoC_i + \beta_{20} RM.MS_i + \beta_{21} SC_i \\ & + \beta_{22} HC_i + \beta_{23} NoT_i + \in_i \end{aligned} \tag{4}$$

where, $N.QDAP_i$ is the score for the negative QDAP sentiment polarity and $P.QDAP_i$ is the score for the positive QDAP sentiment polarity.

4. Results

4.1. Which country is the leading country in terms of raised funding or number of launched ICOs?

We observe from Figure 2. that the U.S. acquired the largest amount of raised funding and is at the same time the leading country in terms of number of launched ICOs. This in an interesting finding because in the IPO market a somewhat reverse picture was recently presented. For instance, in 2019, there were 404 IPOs in China but only 232 in the U.S. We clustered ICOs into 20 distinct industries. In this regard, it is noteworthy that one ICO can be in several industry sectors. From Figure 2. we observe that more than half of the ICOs produce products related to four sectors, which are platforms, business and services, cryptocurrencies and big data. In Table 1. we report the variables that we were able to identify using the information provided in whitepapers in association with various additional internet websites as illustrated in Figure 1.

4.2. Descriptive statistics

In Table 2. we report the descriptive statistics of our variables. For instance, from Table 2. we observe that based on N=1,507 whitepapers used in our analysis, a whitepaper has on average 4,338 words. The minimum number of words is 241 and the maximum number of words is 41,532. Next, let us consider the number of categories. We observe that the average ICO's product is associated with 1.65 categories. The minimum number of categories is one, whereas the maximum number of categories is as much as eleven. The raised amount shows a very high standard deviation among ICOs, where the minimum raised amount is 50 dollars and the maximum amount corresponds to 4 billion dollars. The average funding amount is 20 million dollars. However, we note that the distribution of the raised funding is highly skewed and some few projects raised billions of dollars[10].

4.3. What do we learn from analyzing our regression models?

4.3.1. Does sentiment have an impact on the success of ICOs?

We start our statistical analysis using a simple OLS regression incorporating the sentiment measured by the Harvard Sentiment Dictionary. The results are reported in Table 3. and show some interesting findings. First and most importantly, we find that only negative sentiment is significant. Specifically, the more negative the sentiment the larger the predicted amount of raised funding, whereas positive sentiment does not have any significant effects.

Furthermore, to get a deeper understanding on what emotion is mostly associated with negative or positive sentiment we followed the NRC Word-Emotion Association Lexicon Mohammad and Turney (2013) which is also known as EmoLex[11]. It is a list of English words and their associations with eight basic emotions, which are anger, fear, anticipation, trust, surprise, sadness, joy, and disgust. Using EmoLex also enables us to capture negative and positive sentiments in the text. Figure 6. shows a histogram and a pie chart of sentiment and its corresponding emotion in the whitepapers of our data sample. We observe from Figure 6. that 57% of the overall text in the whitepaper has negative sentiment. Of this negative sentiment, fear is the mostly frequently identified emotion associated with the whitepapers. This might be the indication of an ICO marketing strategy, where customers are attracted by the trigger of fear. We further explored all the words associated with 'fear' in the NRC Lexicon and track them in the corpus of whitepapers and find that the ICO whitepapers (2014-2019) are selling '*fear of risk*', '*fear of change*', '*fear of problem*', '*fear of regulation*', '*fear of loss*' and other fears that triggers the negative emotions.

4.3.2. Is it important whether or not the whitepaper content is easy to read?

The regression result reported in Tables 4-6. incorporates seven different models with seven different readability scores. Our findings strongly suggest that readability is irrelevant irrespective of how we measure it, whereas negative Harvard psychosocial sentiment is statistically significant across all model specifications.

Another interesting and somewhat surprising finding is that neither the number of words, nor the length or the similarity are associated with raised funding. This suggests that investors do not critically review the whitepaper. If investors paid more attention to scrutinizing whitepapers, the risk of deception could be decreased. This is an important issue because the vast majority of ICOs are scam.[12] Our results also show that the more social media attention an ICO receives, the higher the predicted amount of raised funding. This is another issue which may suggest that ICO investors are guided by emotional experiences rather than critical reasoning. This view may be additionally substantiated by the insignificance of the risk scores, as risk scores do not have an impact on the amount of raised funding. Risky investments should be priced differently from less risky investments, but we do not find evidence to support this in our study. One important finding about social media followers, which is captured by the variable 'Hype Score', shows that not rated

[10] The maximum amount of raised funding is $4.2 billion.

[11] see more at https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

[12] The Satis Research Group report of 11 July 2018 investigated approximately 1,500 ICOs whereof 78% were *identified scams*, corresponding to a monetary equivalent in terms of US dollar of $1.3B. On the other hand, slightly more than $8B (~70% of ICO fundraising) was allocated to those that moved on to trade on an exchange. Even though the vast majority of funding was funneled to ICOs that proceeded to trade, about 1,170 out of 1,500 projects were revealed as fraud. The most well-known ICO scams are Pincoin, Arisebank and Savedroid, that illicitly obtained $660M, $600M, and $50M, respectively.

Table 3
Regression Result with Harvard Sentiment Dictionary (GI), Readability and Other ICO Characteristics

	Dependent variable: ln(Raised)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Word Count	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002
	(0.619)	(0.623)	(0.610)	(0.642)	(0.631)	(0.613)	(0.612)
Page Lengths	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002
	(-0.461)	(-0.448)	(-0.458)	(-0.455)	(-0.448)	(-0.441)	(-0.473)
Negative.GI Sentiment	6.856**	6.741**	6.870**	6.800**	6.746**	6.764**	6.920**
	(2.522)	(2.481)	(2.534)	(2.507)	(2.485)	(2.497)	(2.556)
Positive.GI Sentiment	0.523	0.443	0.520	0.535	0.453	0.450	0.565
	(0.439)	(0.371)	(0.437)	(0.449)	(0.380)	(0.377)	(0.463)
Flesch Readbility Score	0.001						
	(0.307)						
Flesch.K Readbility Score		-0.009					
		(-0.665)					
RIX Readbility Score			-0.006				
			(-0.336)				
SMOG Readbility Score				-0.013			
				(-0.547)			
FOG Readbility Score					-0.009		
					(-0.681)		
ARI Readbility Score						-0.008	
						(-0.752)	
Col Readbility Score							0.001
							(0.181)
Jaccard Similarity Score	0.472	0.475	0.472	0.475	0.476	0.476	0.472
	(1.350)	(1.359)	(1.351)	(1.357)	(1.361)	(1.360)	(1.348)
High Hype Score	1.030***	1.032***	1.030***	1.030***	1.032***	1.032***	1.028***
	(3.498)	(3.504)	(3.499)	(3.498)	(3.504)	(3.506)	(3.491)
Medium Hype Score	0.754***	0.758***	0.755***	0.755***	0.758***	0.758***	0.752***
	(2.695)	(2.707)	(2.697)	(2.697)	(2.707)	(2.709)	(2.687)
Low Hype Score	0.184	0.185	0.184	0.183	0.185	0.186	0.183
	(0.664)	(0.666)	(0.664)	(0.661)	(0.667)	(0.670)	(0.661)
Hype Not Rated	0.637**	0.640**	0.638**	0.637**	0.640**	0.641**	0.636**
	(2.274)	(2.282)	(2.276)	(2.272)	(2.283)	(2.286)	(2.268)
High Risk Score	-0.175	-0.169	-0.169	-0.163	-0.168	-0.162	-0.178
	(-0.304)	(-0.293)	(-0.294)	(-0.282)	(-0.291)	(-0.280)	(-0.309)
Medium Risk Score	-0.340	-0.337	-0.335	-0.329	-0.337	-0.331	-0.340
	(-0.601)	(-0.598)	(-0.593)	(-0.583)	(-0.597)	(-0.586)	(-0.602)
Low Risk Score	0.046	0.048	0.051	0.056	0.047	0.054	0.045
	(0.078)	(0.082)	(0.087)	(0.095)	(0.081)	(0.093)	(0.077)
Risk Not Rated	-0.319	-0.315	-0.314	-0.308	-0.316	-0.309	-0.321
	(-0.567)	(-0.561)	(-0.558)	(-0.548)	(-0.562)	(-0.549)	(-0.570)
Number of Categories	-0.009	-0.009	-0.009	-0.009	-0.009	-0.009	-0.008
	(-0.228)	(-0.237)	(-0.225)	(-0.233)	(-0.237)	(-0.242)	(-0.219)
Team.Size	0.013*	0.013*	0.013*	0.013*	0.013*	0.013*	0.013*
	(1.869)	(1.863)	(1.870)	(1.867)	(1.862)	(1.865)	(1.872)
County Disclosed	-0.214	-0.219	-0.214	-0.216	-0.219	-0.219	-0.211
	(-1.147)	(-1.174)	(-1.149)	(-1.161)	(-1.174)	(-1.177)	(-1.133)
KYC Score	0.104**	0.106**	0.104**	0.105**	0.106**	0.106**	0.103*
	(1.976)	(2.005)	(1.979)	(1.992)	(2.005)	(2.007)	(1.961)
Twitter Followers	0.00001**	0.00001**	0.00001**	0.00001**	0.00001**	0.00001**	0.00001**
	(2.347)	(2.368)	(2.352)	(2.348)	(2.369)	(2.376)	(2.333)
RoadMap/Milestone Stated	0.071	0.072	0.071	0.073	0.072	0.072	0.071
	(0.711)	(0.716)	(0.709)	(0.726)	(0.721)	(0.714)	(0.707)
SoftCap Given	-0.243	-0.241	-0.243	-0.242	-0.242	-0.241	-0.244
	(-1.577)	(-1.564)	(-1.573)	(-1.566)	(-1.567)	(-1.560)	(-1.585)
HardCap Given	0.098	0.098	0.097	0.098	0.098	0.097	0.097
	(0.671)	(0.672)	(0.668)	(0.676)	(0.673)	(0.669)	(0.669)
Number of Tokens Given	0.165	0.163	0.165	0.164	0.163	0.163	0.166
	(1.374)	(1.359)	(1.372)	(1.364)	(1.360)	(1.357)	(1.384)
Constant	14.428***	14.617***	14.509***	14.660***	14.638***	14.594***	14.414***
	(20.742)	(20.502)	(21.090)	(19.276)	(20.300)	(20.900)	(19.088)
N	1,507	1,507	1,507	1,507	1,507	1,507	1,507
R^2	0.056	0.056	0.056	0.056	0.056	0.056	0.056
Adjusted R^2	0.041	0.042	0.041	0.042	0.042	0.042	0.041
Residual Std. Error (df = 1483)	1.745	1.744	1.745	1.745	1.744	1.744	1.745
F Statistic (df = 23; 1483)	3.832***	3.848***	3.833***	3.841***	3.849***	3.854***	3.829***

This table reports the OLS regression result with 7 different readability measures in the model (1) to (7) with Flesch, Flesch. K, RIX, SMOG, FOG, ARI, CoL readability scores respectively, incorporating the Harvard General Psychology Sentiment Dictionary. Negative.GI Sentiment is the negative psychological sentiment and Positive.GI Sentiment is the positive psychological sentiment. This table also accounts for social media and bitcoin talk forum hype as High, Medium, and Low Hype Score. Potential fraud risk is measured as High, Medium, and Low-Risk Score. Teams, Twitter followers, Token Economics and other whitepaper disclosures are the other explanatory variables

Note: Significance Levels

t-stats reported in the parenthesis.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

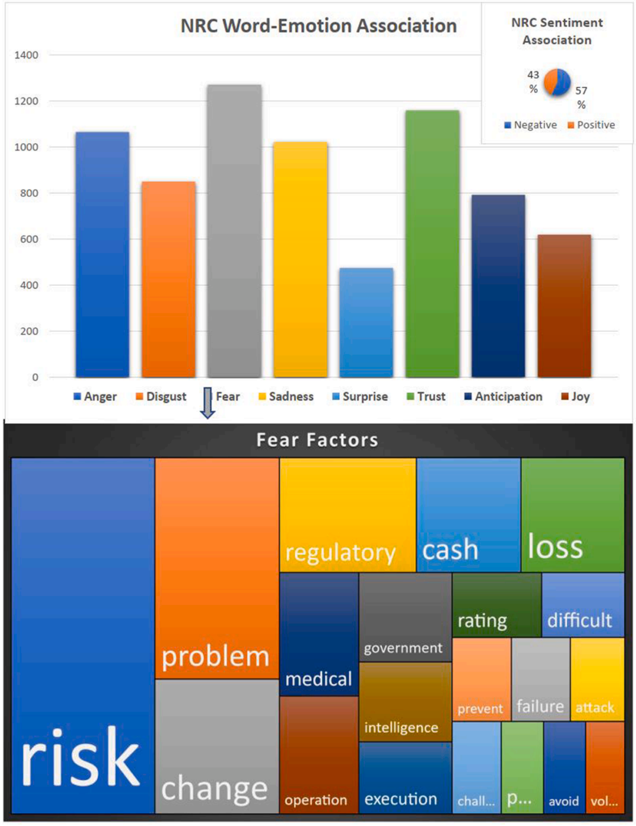


Fig. 6. Unique Words of ICO Whitepapers (2014-2019) and NRC Word-Emotion Association

ICOs tend to be more successful in raising funds than those with a lower number of social media followers. Specifically, ICOs with no social media account are more successful in raising funds than ICOs with very little social media followers. Investors may perceive the quantity of social media followers as a proxy for the popularity of a project. A project with a smaller number of followers may imply that the project is less popular among other investors. Therefore, it is important for the ICOs to have excellent social media marketing strategies from the very beginning. On the other hand, avoiding usage of social media channels for marketing an ICO project may result in investors searching for other factors, such as *team members*, to assess the quality of the project.

Considering that Howell el al. (2020) document that ICO success is associated with disclosure, credible commitment to the project, and quality signals, it is surprising that our evidence gives a somewhat mixed picture: Specifically, county disclosure, road map or explicitly stated softcap or hardcap are all quality signals but our findings indicate that none of these is associated with ICO success in terms of raised funding. On the other hand, it is not surprising that our results show that twitter followers and signature campaigns, as measured by the KYC score, have an effect. Both variables are positively correlated with raised funding. Again, our results suggest that ICO investors base their decisions on attention signals and are attracted to ICOs that are frequently advertised in social media.

Next, we use the sentiment measured by Henry's Business Communication Dictionary, Loughran Mc Donald Finance-specific Dictionary, and the QDAP Sentiment Polarity Dictionary. For each mode, we use various readability measures. The results are reported in Table 5**.**, Table 6**., and** Table 7**.** The main difference between the results reported in Table 5**.**, Table 6**., and** Table 7**.** as opposed to those reported in Table 3**.** is that the sentiment measured by those sentiment dictionaries is statistically insignificant. However, the statistical significances of all those other variables as discussed earlier do not change and, impressively, the point

Table 4
Regression Result with Henry's Finance Dictionary (HE), Readability and Other ICO Characteristics

	Dependent variable: ln(Raised)						
	(1)	(2)	(3)	(4)	(5)	**(6)**	**(7)**
Word Count	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002
	(0.785)	(0.785)	(0.771)	(0.812)	(0.796)	(0.770)	(0.773)
Page Lengths	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.003
	(-0.605	(-0.587)	(-0.606)	(-0.604)	(-0.588)	(-0.584)	(-0.640)
Negative.HE Sentiment	4.209	3.630	3.900	4.401	3.899	3.783	4.095
	(0.338)	(0.292)	(0.314)	(0.353)	(0.314)	(0.304)	(0.328)
Positive.HE Sentiment	3.688	3.358	3.818	3.762	3.395	3.454	3.896
	(0.598)	(0.544)	(0.620)	(0.611)	(0.550)	(0.560)	(0.631)
Flesch Readbility Score	0.003						
	(0.656)						
Flesch.K Readbility Score		-0.014					
		(-1.091)					
RIX Readbility Score			-0.011				
			(-0.627)				
SMOG Readbility Score				-0.020			
				(-0.844)			
FOG Readbility Score					-0.014		
					(-1.084)		
ARI Readbility Score						-0.011	
						(-1.096)	
Col Readbility Score							0.002
							(0.209)
Jaccard Similarity Score	0.468	0.473	0.468	0.471	0.474	0.472	0.466
	(1.333)	(1.349)	(1.335)	(1.342)	(1.350)	(1.347)	(1.329)
High Hype Score	1.044***	1.046***	1.045***	1.043***	1.046***	1.046***	1.041***
	(3.537)	(3.543)	(3.540)	(3.534)	(3.543)	(3.546)	(3.526)
Medium Hype Score	0.777***	0.781***	0.778***	0.776***	0.780***	0.781***	0.773***
	(2.769)	(2.784)	(2.772)	(2.768)	(2.783)	(2.784)	(2.756)
Low Hype Score	0.194	0.194	0.194	0.192	0.195	0.196	0.193
	(0.699)	(0.700)	(0.700)	(0.693)	(0.701)	(0.705)	(0.696)
Hype Not Rated	0.666**	0.668**	0.667**	0.665**	0.669**	0.670**	0.665**
	(2.371)	(2.381)	(2.377)	(2.367)	(2.382)	(2.385)	(2.367)
High Risk Score	-0.171	-0.165	-0.162	-0.155	-0.164	-0.156	-0.179
	(-0.297)	(-0.286)	(-0.280)	(-0.268)	(-0.284)	(-0.271)	(-0.311)
Medium Risk Score	-0.334	-0.334	-0.326	-0.320	-0.334	-0.325	-0.338
	(-0.590)	(-0.590)	(-0.576)	(-0.566)	(-0.590)	(-0.575)	(-0.597)
Low Risk Score	0.040	0.040	0.048	0.053	0.039	0.049	0.036
	(0.068)	(0.069)	(0.082)	(0.091)	(0.068)	(0.084)	(0.062)
Risk Not Rated	-0.324	-0.322	-0.317	-0.311	-0.323	-0.313	-0.331
	(-0.576)	(-0.571)	(-0.561)	(-0.551)	(-0.574)	(-0.556)	(-0.586)
Number of Categories	-0.008	-0.009	-0.008	-0.008	-0.008	-0.009	-0.007
	(-0.210)	(-0.220)	(-0.200)	(-0.209)	(-0.219)	(-0.222)	(-0.184)
Team.Size	0.012*	0.012*	0.012*	0.012*	0.012*	0.012*	0.012*
	(1.770)	(1.765)	(1.772)	(1.768)	(1.763)	(1.768)	(1.771)
County Disclosed	-0.205	-0.212	-0.205	-0.208	-0.212	-0.211	-0.199
	(-1.099)	(-1.134)	(-1.097)	(-1.114)	(-1.134)	(-1.132)	(-1.070)
KYC Score	0.104**	0.106**	0.104**	0.105**	0.106**	0.106**	0.102*
	(1.962)	(2.001)	(1.963)	(1.977)	(1.999)	(1.996)	(1.929)
Twitter Followers	0.00001**	0.00001**	0.00001**	0.00001**	0.00001**	0.00001**	0.00001**
	(2.435)	(2.463)	(2.446)	(2.429)	(2.463)	(2.469)	(2.409)
RoadMap/Milestone Stated	0.072	0.073	0.072	0.074	0.073	0.072	0.071
	(0.720)	(0.723)	(0.714)	(0.737)	(0.730)	(0.718)	(0.706)
SoftCap Given	-0.238	-0.235	-0.237	-0.236	-0.236	-0.235	-0.241
	(-1.540)	(-1.520)	(-1.536)	(-1.529)	(-1.525)	(-1.518)	(-1.558)
HardCap Given	0.106	0.106	0.105	0.107	0.106	0.105	0.106
	(0.729)	(0.725)	(0.722)	(0.733)	(0.726)	(0.720)	(0.724)
Number of Tokens Given	0.149	0.147	0.149	0.148	0.147	0.147	0.152
	(1.241)	(1.222)	(1.239)	(1.234)	(1.225)	(1.224)	(1.260)
Constant	14.812***	15.116***	14.976***	15.195***	15.145***	15.065***	14.857***
	(22.123)	(22.436	(22.840)	(20.898)	(22.212)	(22.735)	(20.775)
N	1,507	1,507	1,507	1,507	1,507	1,507	1,507
R^2	0.051	0.052	0.051	0.052	0.052	0.052	0.051
Adjusted R^2	0.037	0.037	0.037	0.037	0.037	0.037	0.036
Residual Std. Error (df = 1483)	1.749	1.749	1.749	1.749	1.749	1.749	1.749
F Statistic (df = 23; 1483)	3.490***	3.524***	3.488***	3.502***	3.524***	3.525***	3.472***

This table reports the OLS regression result with 7 different readability measures in the model (1) to (7) with Flesch, Flesch. K, RIX, SMOG, FOG, ARI, CoL readability scores respectively, incorporating the Henry's Finance Sentiment Dictionary. Negative.HE Sentiment is the negative financial sentiment and Positive.HE Sentiment is the positive financial sentiment. This table also accounts for social media and bitcoin talk forum hype as High, Medium, and Low Hype Score. Potential fraud risk is measured as High, Medium, and Low-Risk Score. Teams, Twitter followers, Token Economics and other whitepaper disclosures are the other explanatory variables.
Significance Levels
t-stats reported in the parenthesis.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

estimates are virtually the same. This result suggests that the sentiment incorporated in whitepapers cannot be cached by Henry's Business Communication Dictionary, Loughran Mc Donald Finance-specific Dictionary or the QDAP Sentiment Polarity Dictionary. This is an interesting issue because the Loughran Mc Donald Finance-specific Dictionary has been exclusively created because of the inability of standard sentiment dictionaries to measure sentiment in finance-specific contexts.

4.4. Additional robustness check implementing Artificial Neural Networks (ANN).

Textural analysis is either Lexicon-based or Machine Learning based.[13] As a main tool we followed Lexicon-based (i.e. Readability Measure and Dictionary-based). The Machine Learning based textural analysis includes Naïve Bayes, Support Vector Machine, Semantic Analysis and, Neural Network. Recently, Artificial Neural Network (ANN) is gaining more attention in the field of big data and machine learning. ANN is becoming the first choice for the researcher who is stepping into the field of Deep Learning. As an additional robustness check, we also implemented ANN to see which sentiment dictionary is the best fit for the linear model. We followed the min-max normalization method to scale our dataset. Excepts for the dummy variables and negative sentiment (*although they are negative sentiments but the sentiments weights are non-negative values*) and positive sentiment scores under each dictionary, we also scaled *Word Count, Page Length, Number of Categories, Team Size, Twitter Followers* and *Raised Amount*. In addition to these variables we decided to use the *RIX, SMOG,* and *FOG* readability scores as they have non-negative values (see Table 2**.**). The *min-max* normalization process scales the variables between 0 and 1 which feeds machine uniform sets variables in the dataset.

$$x_{scaled} = \frac{x - min(x)}{max(x) - min(x)} \tag{5}$$

where, x represents a single feature or a variable vector, $min(x)$ is the minimum value amongst variable vector x, and $max(x)$ is the maximum amongst variable vector x.

We apply a neural network package (*neuralnet*) available in R statistical software[14]. Furthermore, the calculation of generalized weights is implemented. The data frame is amended by a mean response, the mean of all responses corresponding to the same covariate-vector. To obtain an overview of the results of the neural network and the generalized linear model objects, the covariate matrix is bound to the output of the neural network and the fitted values of the generalized linear model object.

We begin the ANN with 75% training and 25% test data implementing 3 hidden layers without any hidden neurons to each model incorporating different sentiment dictionaries and one readability score RIX (see **Appendix A.9. – A.12.**). The implemented neural network setup gives the lowest mean squared errors (MSE) to the test data that incorporates the Harvard GI sentiment dictionary in the linear model, the MSE on different ANN setups are given in panel a of Table 7**.** Similarly, Figure 7**.** shows the ANN model plot fitting on four different sentiment dictionaries with 75% training and 25% test data with 3 hidden layers. QDAP dictionary also has an equivalent MSE value, whereas Henry's and Loughran McDonald finance specific dictionaries have the first and the second highest MSE among all four. This might be the indication that finance specific sentiment dictionaries do not accurately capture fintech specific sentiments.

The predictive power of ANN is subject to change based on the proportion of training and test data as well as the number of hidden layers and neurons. Keeping the training and testing proportion same as previous, this time we increased the hidden layers to 10 from 3 and added 5 neurons to the ANN setup. Surprisingly, each model incorporating different sentiment dictionaries and RIX readability score gives the same mean squared errors to all the models for the test data. The fitted model plot for the test data is given in Figure 8.

Implementing ANN with higher number of hidden layers makes it difficult to decide on which sentiment dictionary best suits the model as mean squared errors are lower and equal for all. Instead of adding more hidden layers and hidden neurons, we increased the proportion of training data to 90%. Adding three hidden layers and no hidden neurons, again the liner model implementing ANN gives the lowest mean squared errors to the test data with model that incorporates Harvard GI sentiment dictionary. Figure 9**.** shows the ANN model plot fitting on four different sentiment dictionaries with 90% training and 10% test data with 3 hidden layers. This time, QDAP and Loughran McDonald dictionaries got the same accuracy, whereas Henry's finance specific dictionary is still giving the highest MSE value. Furthermore, we again increased the hidden layers to 10 from 3 and added 5 hidden neurons, the test results in Figure 10**.** show

[13] How the choices of our approach are in line with Textural Analysis is illustrated in **Appendix A.8.**

[14] This package utilizes the training of neural networks using the backpropagation, resilient backpropagation with or without weight backtracking. The package allows flexible settings through custom-choice of error and activation function. See more at; https://cran.r-project.org/web/packages/neuralnet/neuralnet.pdf

Table 5
Regression Model with Loughran Mc Donald Finance-specific Dictionary (LM), Readability and Other ICO Characteristics.

	Dependent variable: ln(Raised)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Word Count	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002
	(0.797)	(0.796)	(0.782)	(0.822)	(0.806)	(0.781)	(0.786)
Page Lengths	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003	-0.003
	(-0.655)	(-0.642)	(-0.655)	(-0.654)	(-0.642)	(-0.637)	(-0.686)
Negative.LM Sentiment	-2.648	-2.840	-2.586	-2.717	-2.841	-2.766	-2.391
	(-0.778)	(-0.833)	(-0.761)	(-0.798)	(-0.833)	(-0.813)	(-0.706)
Positive.LM Sentiment	5.638	5.285	5.666	5.642	5.318	5.372	5.827
	(1.203)	(1.125)	(1.209)	(1.204)	(1.132)	(1.145)	(1.234)
Uncertain.LM Sentiment	3.845	3.615	3.708	4.074	3.810	3.650	3.750
	(0.488)	(0.459)	(0.471)	(0.516)	(0.484)	(0.464)	(0.476)
Flesch Readbility Score	0.003						
	(0.702)						
Flesch.K Readbility Score		-0.015					
		(-1.109)					
RIX Readbility Score			-0.012				
			(-0.660)				
SMOG Readbility Score				-0.021			
				(-0.886)			
FOG Readbility Score					-0.014		
					(-1.105)		
ARI Readbility Score						-0.011	
						(-1.110)	
Col Readbility Score							0.002
							(0.283)
Jaccard Similarity Score	0.435	0.442	0.436	0.438	0.442	0.441	0.434
	(1.239)	(1.257)	(1.240)	(1.247)	(1.258)	(1.255)	(1.234)
High Hype Score	1.049***	1.051***	1.049***	1.048***	1.051***	1.052***	1.044***
	(3.550)	(3.559)	(3.552)	(3.547)	(3.559)	(3.561)	(3.534)
Medium Hype Score	0.787***	0.792***	0.788***	0.787***	0.792***	0.792***	0.782***
	(2.805)	(2.822)	(2.807)	(2.804)	(2.821)	(2.821)	(2.786)
Low Hype Score	0.205	0.206	0.205	0.204	0.206	0.207	0.203
	(0.739)	(0.742)	(0.739)	(0.734)	(0.743)	(0.746)	(0.732)
Hype Not Rated	0.678**	0.681**	0.679**	0.677**	0.681**	0.682**	0.675**
	(2.411)	(2.423)	(2.416)	(2.409)	(2.424)	(2.426)	(2.402)
High Risk Score	-0.182	-0.174	-0.173	-0.165	-0.173	-0.166	-0.191
	(-0.315)	(-0.302)	(-0.299)	(-0.285)	(-0.300)	(-0.288)	(-0.331)
Medium Risk Score	-0.342	-0.341	-0.334	-0.327	-0.340	-0.333	-0.346
	(-0.604)	(-0.602)	(-0.590)	(-0.577)	(-0.602)	(-0.587)	(-0.611)
Low Risk Score	0.034	0.035	0.042	0.048	0.035	0.044	0.029
	(0.057)	(0.060)	(0.071)	(0.082)	(0.059)	(0.075)	(0.050)
Risk Not Rated	-0.334	-0.331	-0.327	-0.320	-0.332	-0.323	-0.341
	(-0.593)	(-0.587)	(-0.579)	(-0.567)	(-0.589)	(-0.573)	(-0.605)
Number of Categories	-0.009	-0.009	-0.008	-0.009	-0.009	-0.009	-0.008
	(-0.229)	(-0.238)	(-0.219)	(-0.226)	(-0.236)	(-0.240)	(-0.204)
Team.Size	0.012*	0.012*	0.012*	0.012*	0.012*	0.012*	0.012*
	(1.773)	(1.766)	(1.775)	(1.771)	(1.764)	(1.770)	(1.776)
County Disclosed	-0.195	-0.202	-0.195	-0.198	-0.202	-0.201	-0.189
	(-1.045)	(-1.080)	(-1.043)	(-1.059)	(-1.080)	(-1.077)	(-1.015)
KYC Score	0.101*	0.103*	0.101*	0.102*	0.103*	0.103*	0.099*
	(1.914)	(1.952)	(1.916)	(1.930)	(1.951)	(1.947)	(1.880)
Twitter Followers	0.00001**	0.00001**	0.00001**	0.00001**	0.00001**	0.00001**	0.00001**
	(2.472)	(2.501)	(2.481)	(2.466)	(2.501)	(2.506)	(2.440)
RoadMap/Milestone Stated	0.072	0.072	0.071	0.074	0.073	0.072	0.070
	(0.713)	(0.717)	(0.707)	(0.731)	(0.724)	(0.712)	(0.699)
SoftCap Given	-0.231	-0.228	-0.230	-0.229	-0.229	-0.228	-0.233
	(-1.490)	(-1.476)	(-1.486)	(-1.478)	(-1.479)	(-1.473)	(-1.506)
HardCap Given	0.100	0.100	0.099	0.101	0.100	0.099	0.099
	(0.685)	(0.684)	(0.678)	(0.687)	(0.684)	(0.679)	(0.679)
Number of Tokens Given	0.148	0.146	0.147	0.147	0.146	0.146	0.150
	(1.227)	(1.210)	(1.225)	(1.220)	(1.212)	(1.211)	(1.247)
Constant	14.800***	15.129***	14.975***	15.210***	15.159***	15.072***	14.820***
	(21.876)	(21.956)	(22.435)	(20.483)	(21.733)	(22.301)	(20.392)
N	1,507	1,507	1,507	1,507	1,507	1,507	1,507
R^2	0.052	0.053	0.052	0.052	0.053	0.053	0.052
Adjusted R^2	0.037	0.037	0.037	0.037	0.037	0.037	0.037
Residual Std. Error (df = 1482)	1.749	1.748	1.749	1.749	1.748	1.748	1.749
F Statistic (df = 24; 1482)	3.400***	3.433***	3.398***	3.413***	3.432***	3.433***	3.382***

This table reports the OLS regression result with 7 different readability measures in the model (1) to (7) with Flesch, Flesch. K, RIX, SMOG, FOG, ARI, CoL readability scores respectively, incorporating the Loughran Mc Donald Finance-specific Dictionary. Negative.LM Sentiment is the negative financial sentiment, Positive.LM Sentiment is the positive financial sentiment, and Uncertain.LM Sentiment is the uncertainty in the sentiment. This table also accounts for social media and bitcoin talk forum hype as High, Medium, and Low Hype Score. Potential fraud risk is measured as High, Medium, and Low-Risk Score. Teams, Twitter followers, Token Economics and other whitepaper disclosures are the other explanatory variables.
Note: Significance Levels
t-stats reported in the parenthesis.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Table 6
Regression Model with QDAP Sentiment Dictionary, Readability and Other ICO Characteristics.

	Dependent variable: *ln(Raised)*						
	(1)	(2)	(3)	(4)	(5)	**(6)**	**(7)**
Word Count	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002	0.00002
	(0.818)	(0.814)	(0.804)	(0.845)	(0.825)	(0.802)	(0.812)
Page Lengths	-0.003	-0.002	-0.003	-0.003	-0.002	-0.002	-0.003
	(-0.612)	(-0.598)	(-0.612)	(-0.610)	(-0.599)	(-0.594)	(-0.640)
Negative.QDAP Sentiment	1.504	1.377	1.459	1.499	1.415	1.409	1.560
	(0.425)	(0.388)	(0.412)	(0.423)	(0.399)	(0.398)	(0.440)
Positive.QDAP Sentiment	2.205	2.052	2.214	2.209	2.070	2.080	2.351
	(1.363)	(1.260)	(1.369)	(1.367)	(1.273)	(1.280)	(1.437)
Flesch Readbility Score	0.002						
	(0.597)						
Flesch.K Readbility Score		-0.012					
		(-0.917)					
RIX Readbility Score			-0.010				
			(-0.556)				
SMOG Readbility Score				-0.019			
				(-0.786)			
FOG Readbility Score					-0.012		
					(-0.928)		
ARI Readbility Score						-0.010	
						(-0.949)	
Col Readbility Score							0.003
							(0.415)
Jaccard Similarity Score	0.466	0.471	0.466	0.469	0.471	0.470	0.465
	(1.330)	(1.343)	(1.331)	(1.339)	(1.345)	(1.342)	(1.327)
High Hype Score	1.033***	1.035***	1.033***	1.032***	1.035***	1.035***	1.029***
	(3.499)	(3.506)	(3.501)	(3.496)	(3.505)	(3.508)	(3.484)
Medium Hype Score	0.764***	0.768***	0.764***	0.763***	0.767***	0.768***	0.759***
	(2.721)	(2.736)	(2.723)	(2.720)	(2.735)	(2.736)	(2.704)
Low Hype Score	0.182	0.183	0.182	0.180	0.183	0.184	0.180
	(0.654)	(0.658)	(0.655)	(0.649)	(0.658)	(0.662)	(0.646)
Hype Not Rated	0.649**	0.652**	0.651**	0.648**	0.652**	0.653**	0.646**
	(2.311)	(2.322)	(2.316)	(2.308)	(2.323)	(2.325)	(2.299)
High Risk Score	-0.178	-0.173	-0.170	-0.163	-0.172	-0.165	-0.183
	(-0.308)	(-0.301)	(-0.295)	(-0.281)	(-0.298)	(-0.286)	(-0.317)
Medium Risk Score	-0.342	-0.342	-0.336	-0.329	-0.342	-0.335	-0.342
	(-0.605)	(-0.606)	(-0.594)	(-0.582)	(-0.605)	(-0.592)	(-0.604)
Low Risk Score	0.028	0.029	0.035	0.041	0.028	0.036	0.028
	(0.048)	(0.049)	(0.060)	(0.070)	(0.048)	(0.062)	(0.048)
Risk Not Rated	-0.330	-0.329	-0.324	-0.317	-0.329	-0.321	-0.332
	(-0.586)	(-0.583)	(-0.574)	(-0.563)	(-0.585)	(-0.570)	(-0.590)
Number of Categories	-0.009	-0.009	-0.009	-0.009	-0.009	-0.010	-0.009
	(-0.237)	(-0.242)	(-0.229)	(-0.237)	(-0.242)	(-0.245)	(-0.224)
Team.Size	0.012*	0.012*	0.012*	0.012*	0.012*	0.012*	0.012*
	(1.785)	(1.780)	(1.787)	(1.783)	(1.778)	(1.783)	(1.788)
County Disclosed	-0.191	-0.197	-0.190	-0.194	-0.197	-0.197	-0.185
	(-1.022)	(-1.054)	(-1.020)	(-1.036)	(-1.054)	(-1.053)	(-0.991)
KYC Score	0.101*	0.103*	0.101*	0.101*	0.102*	0.102*	0.099*
	(1.903)	(1.937)	(1.903)	(1.918)	(1.936)	(1.934)	(1.868)
Twitter Followers	0.00001**	0.00001**	0.00001**	0.00001**	0.00001**	0.00001**	0.00001**
	(2.403)	(2.426)	(2.411)	(2.397)	(2.426)	(2.432)	(2.374)
RoadMap/Milestone Stated	0.072	0.072	0.071	0.074	0.073	0.072	0.071
	(0.716)	(0.718)	(0.710)	(0.733)	(0.724)	(0.715)	(0.707)
SoftCap Given	-0.230	-0.228	-0.229	-0.228	-0.229	-0.228	-0.232

(continued on next page)

Table 6 (*continued*)

	Dependent variable: *ln(Raised)*						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	(-1.490)	(-1.477)	(-1.485)	(-1.479)	(-1.480)	(-1.474)	(-1.503)
HardCap Given	0.101	0.101	0.100	0.102	0.101	0.100	0.100
	(0.693)	(0.691)	(0.686)	(0.696)	(0.692)	(0.687)	(0.689)
Number of Tokens Given	0.148	0.146	0.148	0.147	0.146	0.146	0.149
	(1.228)	(1.215)	(1.227)	(1.221)	(1.216)	(1.215)	(1.243)
Constant	14.575***	14.856***	14.725***	14.931***	14.882***	14.813***	14.533***
	(21.117)	(21.143)	(21.656)	(19.888)	(20.947)	(21.489)	(19.438)
Observations	1,507	1,507	1,507	1,507	1,507	1,507	1,507
R^2	0.053	0.053	0.053	0.053	0.053	0.053	0.052
Adjusted R^2	0.038	0.038	0.038	0.038	0.038	0.038	0.038
Residual Std. Error (df = 1483)	1.748	1.748	1.748	1.748	1.748	1.748	1.748
F Statistic (df = 23; 1483)	3.578***	3.600***	3.576***	3.590***	3.601***	3.603***	3.570***

This table reports the OLS regression result with 7 different readability measures in the model (1) to (7) with Flesch, Flesch. K, RIX, SMOG, FOG, ARI, CoL readability scores respectively, incorporating the Qualitative Discourse Analysis Package Dictionary. Negative.QDAP Sentiment is the negative discourse sentiment and Positive.QDAP Sentiment is the positive discourse sentiment. This table also accounts for social media and bitcoin talk forum hype as High, Medium, and Low Hype Score. Potential fraud risk is measured as High, Medium, and Low-Risk Score. Teams, Twitter followers, Token Economics and other whitepaper disclosures are the other explanatory variables.

Significance Levels

t-stats reported in the parenthesis.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

the lowest MSE for Harvard GI and the highest MSE for QDAP and the same MSE for Henry's and Loughran McDonald dictionaries.

This confirms that the Artificial Neural Network Incorporating Harvard Sentiment Dictionary is a better fitting model. However, the predictive power of the model is subject to change based on different factors like; the number of variables used, the number of hidden layers, the ratio of training and testing data set, etc.

One could argue that the result is subject to change based on the choice of readability score measure. Therefore, we followed the same ANN setups following two other readability measures, SMOG and FOG as they also have non-negative values and same scaling methods can be implemented. We got the lowest and almost the same MSE on average for Harvard GI sentiment dictionary incorporating SMOG and FOG readability measures, which is reported in panel b and panel c in Table 7.. The plot fitting accuracy on four different sentiment dictionaries with SMOG and FOG under different ANN setups are given in **Appendix (A.13- A.20)**. The ANN setups using *RIX*, *SMOG*, and *FOG* also show that our regression model is best fitted with Harvard GI sentiment dictionary, thus the result is robust.

Table 7

Various Artificial Neural Networks setups and Mean Squared Errors (MSE).

ANN	Descriptions (Training%/Test%/HiddenLayers/Neurons)	MSE.GI	MSE.HE	MSE.LM	MSE.QDAP
	Panel a	Readability Index (RIX) Readability Score			
1	75/25/3/0	0.00282	0.00757	0.00413	0.00284
2	90/10/3/0	0.00025	0.00029	0.00026	0.00026
3	75/25/10/5	0.00280	0.00280	0.00280	0.00280
4	90/10/10/5	0.00025	0.00026	0.00026	0.00027
	Average MSE	**0.00153**	**0.00273**	**0.00186**	**0.00154**
	Panel b	**SMOG Readability Score**			
5	75/25/3/0	0.00280	0.00280	0.00280	0.00280
6	90/10/3/0	0.00025	0.00025	0.00024	0.00024
7	75/25/10/5	0.00280	0.00280	0.00280	0.00290
8	90/10/10/5	0.00023	0.00025	0.00027	0.00025
	Average MSE	**0.00152**	**0.00153**	**0.00153**	**0.00155**
	Panel c	**Gunning Fog (FOG) Readability Score**			
9	75/25/3/0	0.00280	0.00300	0.00280	0.00280
10	90/10/3/0	0.00025	0.00024	0.00024	0.00024
11	75/25/10/5	0.00280	0.00290	0.00290	0.00290
12	90/10/10/5	0.00024	0.00025	0.00027	0.00025
	Average MSE	**0.00152**	**0.00160**	**0.00155**	**0.00155**

This table reports the mean squared errors under various ANN setups implementing three different readability measures incorporating four different sentiment dictionaries and other ICO characteristics.

Note: MSE.GI (Mean Squared Error, Harvard GI Dictionary)

MSE.HE (Mean Squared Error, Henry's Finance Dictionary)

MSE.LM (Mean Squared Error, Loughran Mc Donald Finance-specific Dictionary)

MSE.QDAP (Mean Squared Error, Qualitative Discourse Analysis Package)

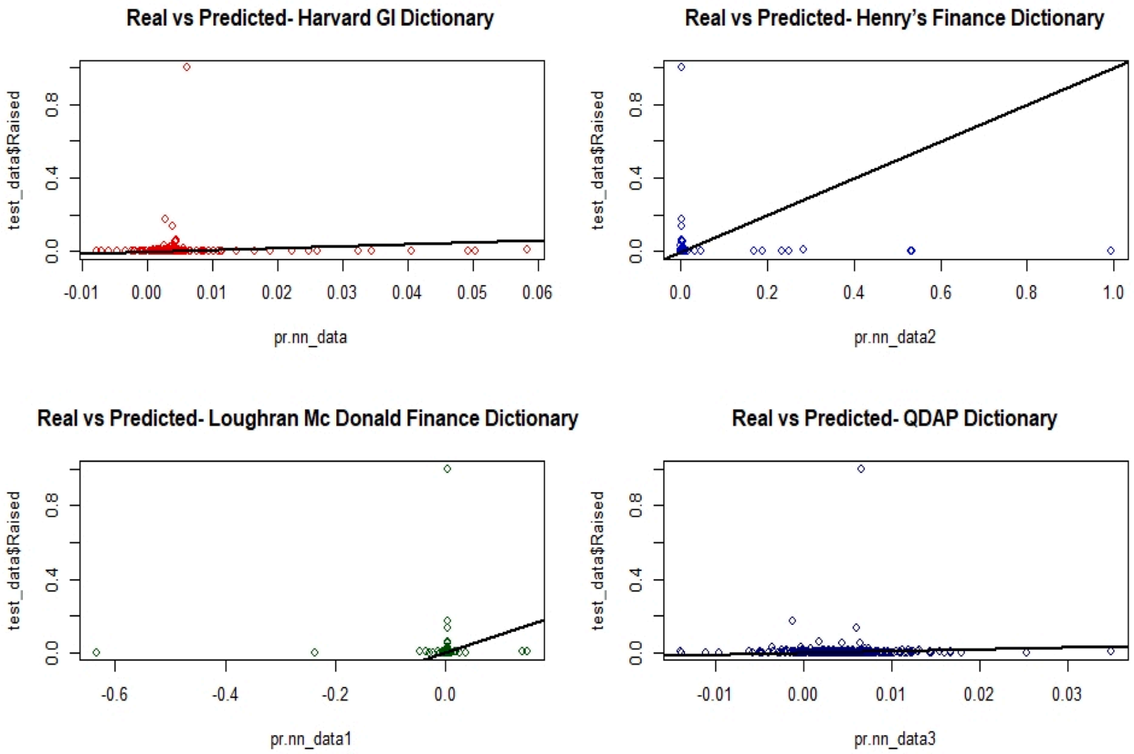


Fig. 7. **ANN Model Fitting on Four Different Sentiment Dictionaries (***75% training and 25% test data with 3 hidden layers, RIX***)**

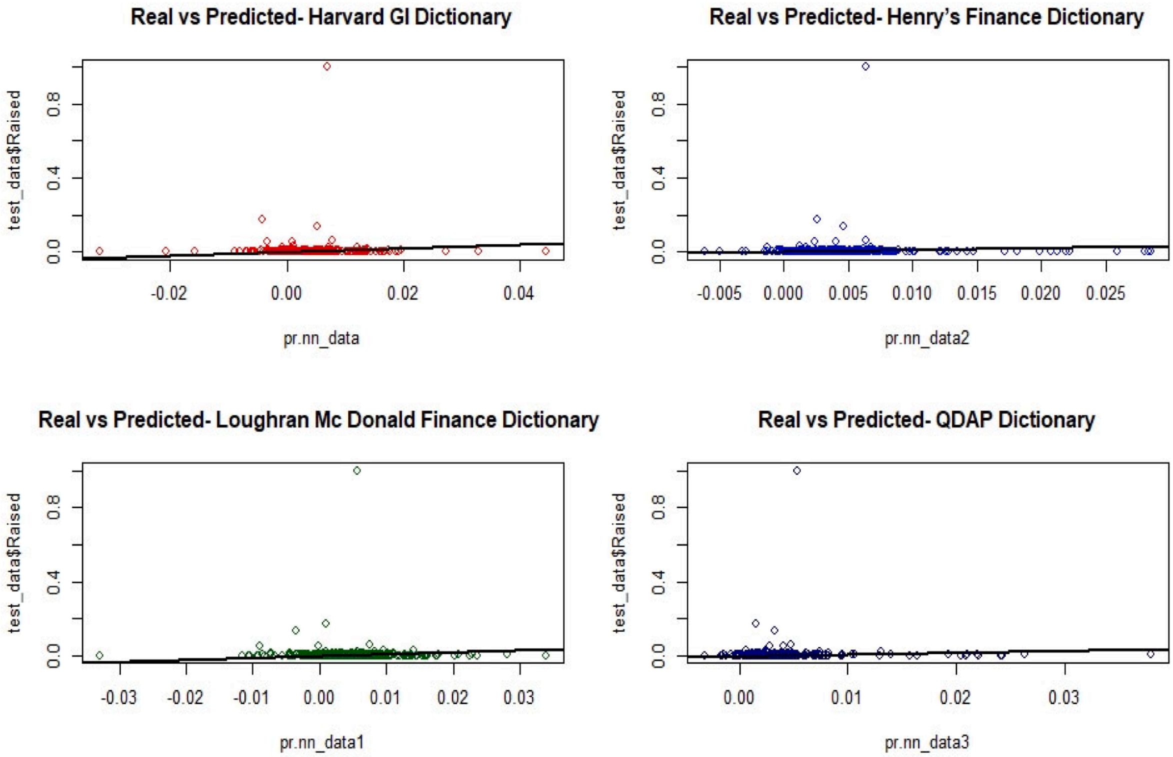


Fig. 8. **ANN Model Fitting on Four Different Sentiment Dictionaries (***75% training and 25% test data with 10 hidden layers, 5 neurons, RIX***)**

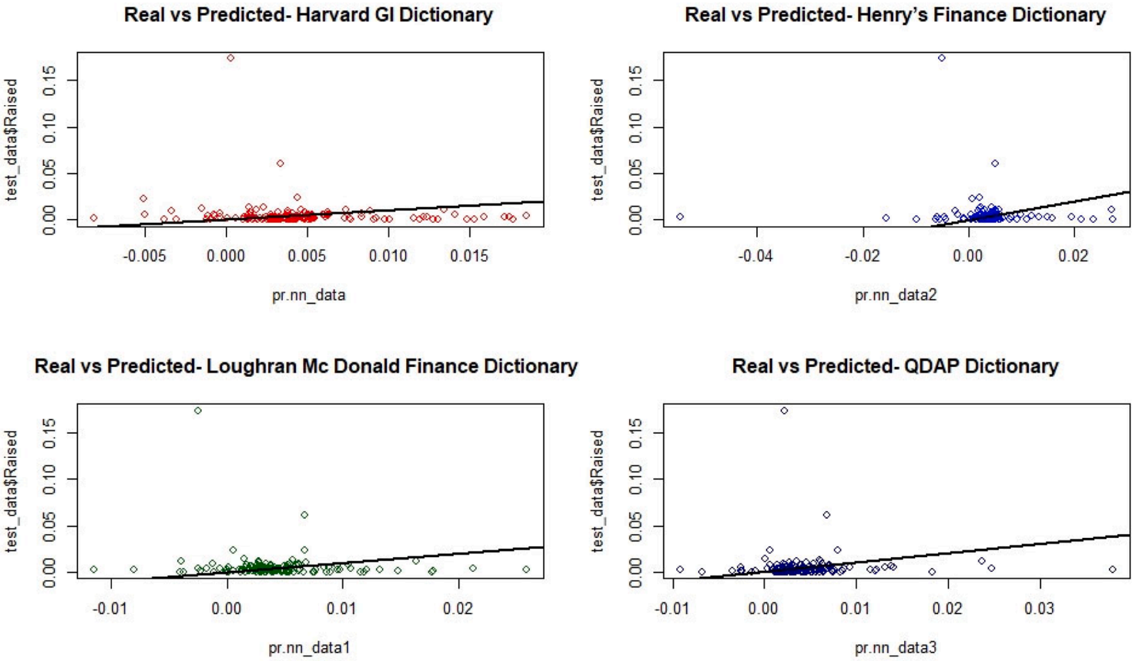


Fig. 9. ANN Model Fitting on Four Different Sentiment Dictionaries (*90% training and 10% test data with 3 hidden layers, and RIX readability score***)**

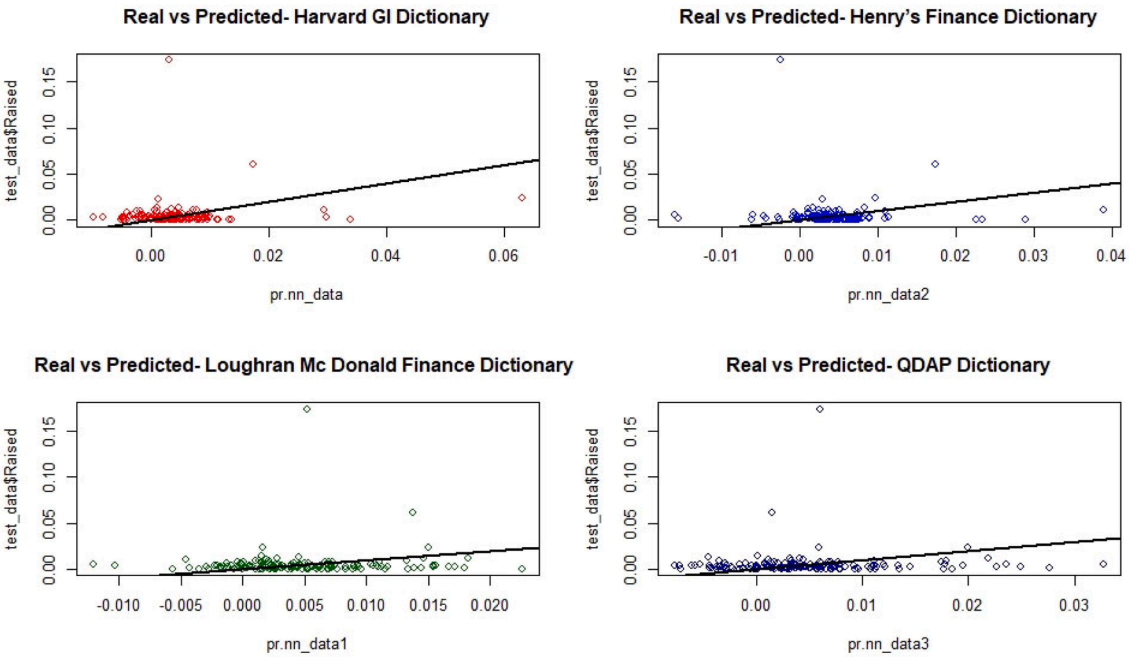


Fig. 10. ANN Model Fitting on Four Different Sentiment Dictionaries (*90% training and 10% test data with 10 hidden layers and 5 neurons, and RIX readability score***)**

5. Conclusion

Extending earlier studies by retrieving the entire population of ICOs that have been launched in the 2014–2019 period, we found 1,507 ICOs that exhibit data on the amount of raised funding. By searching for data on ICO characteristics on various websites in association with textual analysis of those whitepapers, we identified 37 potential variables that could serve as factors associated with the success of ICOs. Contrary to earlier studies, our findings indicate that quality signals such as a number of tokens and/or softcap/hardcap, do not appear to predict ICO success. Also, the readability of a whitepaper, which may serve as an additional indicator of quality, is not associated with ICO success. We hypothesize that a rational investor would intensively deal with an ICO whitepaper and assess a project's quality based on quality and risk assessments. We do not find such evidence either as risk scores are not associated with ICO success.

Interestingly, our results provide strong evidence that ICO investors are mainly guided by their emotional experience when investing in the ICO market. Specifically, we find that negative sentiment in ICO whitepapers is positively associated with the amount of raised funding. This result suggests that negative emotions are an important factor in acquiring funding via ICOs. Moreover, the number of followers on Twitter and the attention that an ICO attract influence ICO success. Specifically, the more followers on Twitter an ICO have the higher the amount of raised funding which may be an indication of herding behavior. Since this behavior is also characterized by a desire to stay continually connected with what others are doing, we argue that the significance of a number of Twitter followers, signature campaigns and the attention scores are clear indications of the significance of this phenomenon. Future research is strongly encouraged to elaborate more on this important issue.

The question arises which type of fear impacts investors' demand for tokens? Our findings indicate that investors' behavior in the ICO market is mainly driven by fears associated to 'risk', 'problem', 'change', and 'regulation', among others. Concerning fear associated with 'risk', for instance, people face nowadays (i) risk of inflation due to extremely low-interest rates in association with quantitative easing, (ii) a risk of global warming due to pollution, (iii) risk of cyberattacks due to lacks in technological advance, among others. Our findings show that projects that successfully communicate in their whitepapers how they address those risks are the successful ICOs in terms of acquiring higher amounts of funding.

Finally, Loughran and McDonald (2011) show that word lists developed for other disciplines misclassify common words in the financial text. However, research shows that the overall sentiment accuracy of Loughran and McDonald dictionary is around 60% even in financial contexts. This indicates that a borrowed dictionary from a different discipline is likely to misjudge the sentiment exponentially. Capturing the true sentiment from the ICO whitepapers plays a significant role in risk management, this paper has important implications for investors willing to finance the project(s) related to blockchain, more specifically by investing in ICOs. Furthermore, analyzing whitepaper sentiments with non-FinTech dictionaries might be one limitation of this study. We could also observe that the sentiment captured by both of the finance-specific dictionaries did not provide any significant results in our analysis. The artificial neural network analysis also favors the Harvard GI psychological sentiment dictionary and confirms that our result is robust. We argue that there can be two different reasons for this phenomenon, that is, either the finance-specific sentiment is of no significance to investors or these dictionaries did not capture the true sentiment in FinTech-related contexts such as ICO. Therefore, we argue that there is an absolute necessity for a FinTech-specific sentiment dictionary that accurately captures the sentiment in the contexts of the new digital financial markets. This is, however, left for future research. ICOs have garnered significant attention in recent years, however, the market is not without its share of scams and fraudulent activity (Grobys, King, & Sapkota, 2022). In a recent study, Chiu et al. (2022) utilized Linguistic Inquiry and Word Count to examine the language used in ICO whitepapers, identifying words and phrases that may indicate potential fraud. Future research may benefit from examining how scammers use whitepaper language to mislead investors, perhaps utilizing a FinTech-specific sentiment dictionary to extract sentiments from these documents.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix

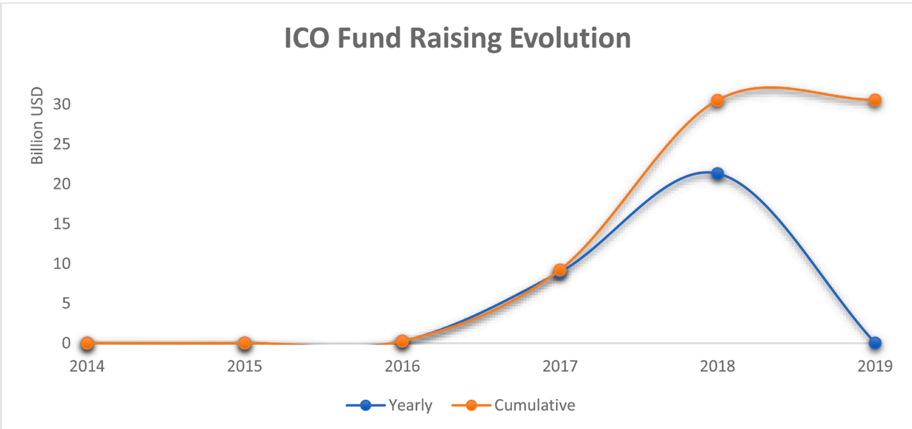
Appendix A.1. R packages and functions and their usage in the paper

S.No.	R Package, function	Usage
1	rvest, xml2	to web scrap the scattered data
2	pdftools, VCorpus	to read pdf files into R and create a corpus
3	tm, tm_map, DocumentTermMatrix, TermDocumentMatrix	cleaning the corpus, creating term and document matrix
4	wordcloud, wordcloud2	to create the word cloud
5	SentimentAnalysis, alalyzesentiment	to get the various sentiment scores of each document corpus
6	textreuse, TextReuseCorpus, pairwise_compare, jaccard_similarity	to get the pairwise Jaccard similarity scores for each document corpus
7	KoRpus, textstat_readability	to get the different readability scores for each document corpus
8	stargazer	to display R outputs into tables
9	neuralnet	to perform artificial neural network

Appendix A.2. Number of ICOs and Amount Raised by Different Countries Around the Globe During 2014-2019 Period.

S. No.	Country	Raised($)	ICOs	S. No.	Country	Raised ($)	ICOs	S. No.	Country	Raised($)	ICOs
1	Non-disclosed	18270321760	765	28	Netherlands	74679825	13	55	Brazil	19629000	1
2	United States	2480961468	110	29	South Korea	70666256	3	56	Ecuador	16314065	1
3	Switzerland	1211340383	49	30	Belize	64416465	9	57	Spain	14959000	5
4	Singapore	1088358207	65	31	Afghanistan	62000000	2	58	Marshall Islands	13044000	2
5	United Kingdom	898694814	59	32	Indonesia	61159000	3	59	Latvia	12648000	6
6	Russia	700688304	58	33	Bulgaria	59031295	7	60	CostaRica	12017000	2
7	Cayman Islands	685258522	13	34	Sweden	49653398	2	61	Panama	11000000	2
8	Estonia	532220166	42	35	Czech Republic	47623000	6	62	Philippines	10864000	1
9	Gibraltar	311165807	15	36	Seychelles	46560000	8	63	Luxembourg	10643815	2
10	HongKong	284419350	20	37	Iceland	45949800	1	64	Cambodia	10561000	2
11	Canada	282268020	17	38	Argentina	44316000	4	65	Bahamas	10000000	1
12	Lithuania	272378963	5	39	Turkey	42868731	3	66	American Samoa	6507000	1
13	Germany	257943033	14	40	Israel	42784000	7	67	Finland	6125000	2
14	Australia	250441332	16	41	Latin America	41710592	1	68	Malaysia	5074000	1
15	Georgia	186652722	9	42	Dubai	41706262	2	69	Norway	3751000	3
16	Japan	186462000	6	43	Thailand	39495517	5	70	Belgium	3609000	1
17	Malta	158261610	12	44	Taiwan	39175060	2	71	Egypt	2877000	1
18	United Arab Emirates	157640000	9	45	Slovakia	37378000	1	72	Serbia	2820000	1
19	France	155191896	10	46	Romania	31513000	3	73	Denmark	1022593	2
20	China	155057000	10	47	Kazakhstan	30280000	2	74	Jersey	1000000	1
21	India	119564600	9	48	Liechtenstein	28682516	2	75	SaintVincent	1000000	1
22	South Africa	112582843	5	49	Austria	28362000	5	76	Italy	708000	2
23	British Virgin Islands	102469132	9	50	Isle of Man	27515417	3	77	Saint Kitts and Nevis	305000	1
24	Poland	101885080	8	51	Mexico	24000000	2	78	Hungary	249000	1
25	Ukraine	88855172	5	52	Belarus	22874000	3	79	Pakistan	51000	1
26	Slovenia	86815000	11	53	Korea	20665996	2	80	Peru	15000	1
27	Cyprus	79269000	8	54	Colombia	40930000	2	81	Andorra	2000	1
									Total	30559957787	1508

Appendix A.3. Evolution of ICO fund raising over time.



Appendix A.**4.** Number of ICOs under each category (same ICO has been listed under various categories)

S.No.	Category	ICOs	S.No.	Category	ICOs
1	Platform	1373	24	Manufacturing	58
2	Business and Services	1260	25	Blockchain	55
3	Cryptocurrency	1015	26	Charity	49
4	Investment	469	27	Payments and Wallets	38
5	Software and Computing	389	28	Legal	33
6	Communication and Media	373	29	Art	33
7	Artificial Intelligence	370	30	Electronics	31
8	Banking and Finance	316	31	Identity and Security	22
9	Internet	306	32	Content	21
10	Smartcontract	295	33	Mining	16
11	Entertainment	285	34	Gaming Industry	16
12	Infrastructure	230	35	Commerce	16
13	Gambling, Casino, Betting	206	36	Marketplace	10
14	Bigdata	174	37	Advertising	9
15	Retail	143	38	Logistics	8
16	Health and Drugs	130	39	Augumented Reality	8
17	Virtual Reality	112	40	Utilities	7
18	Real-Estate	112	41	Jobs	6
19	Education	97	42	Asset Management	6
20	Other	88	43	Venture Capital	5
21	Tourism	87	44	Internet of Things	5
22	Energy	76	45	Funding	5
23	Sports	70	46	Transportation	3

Appendix A.**5.** Top 200 words used in ICO whitepapers (2014-2019)

S.No.	word	freq	S.No.	word	freq	S.No.	word	freq	S.No.	word	freq
1	token	299563	51	provide	32093	101	growth	20656	151	rate	15589
2	use	289016	52	global	32034	102	mining	20179	152	fiat	15571
3	crypto	155546	53	management	31577	103	current	20131	153	internet	15554
4	platform	147976	54	public	31453	104	fund	19801	154	research	15527
5	blockchain	134098	55	future	30518	105	get	19788	155	reward	15383
6	market	105919	56	process	30440	106	open	19686	156	full	15329
7	service	102761	57	experience	30185	107	made	18876	157	features	15264
8	data	102193	58	currency	30167	108	asset	18791	158	take	15135
9	system	89070	59	decentralized	29644	109	year	18605	159	launch	15050
10	network	83274	60	make	29635	110	possible	18550	160	technologies	15027
11	company	81228	61	white	29047	111	sales	18469	161	share	15005
12	transaction	80592	62	amount	28825	112	potential	18414	162	proof	14859
13	contract	77877	63	available	28764	113	applications	18383	163	start	14838
14	exchange	75775	64	purchase	28084	114	receive	18351	164	capital	14831
15	pay	72459	65	chain	27543	115	level	18264	165	help	14736
16	project	69882	66	high	27320	116	case	18052	166	parties	14724
17	innovation	65650	67	legal	27187	117	game	17929	167	required	14714
18	information	63264	68	order	27167	118	participants	17926	168	ensure	14706
19	business	62271	69	investors	26836	119	usd	17840	169	additional	14659
20	time	61990	70	real	26805	120	limited	17776	170	means	14567
21	ico	61572	71	people	26617	121	terms	17674	171	provides	14419
22	smart	60384	72	bitcrypto	25895	122	supply	17409	172	trust	14392
23	development	60287	73	online	25621	123	holders	17258	173	control	14377
24	eth	58331	74	create	25338	124	allows	17182	174	offering	14359
25	technology	58321	75	work	25245	125	party	17166	175	bank	14323
26	sale	55197	76	assets	24588	126	storage	17140	176	document	14320
27	value	53084	77	private	24534	127	media	17102	177	event	14319
28	product	48643	78	total	24285	128	large	17091	178	rewards	14299
29	team	48108	79	key	24241	129	page	16918	179	members	14254
30	risk	44313	80	model	24190	130	rights	16879	180	period	14252
31	world	41952	81	application	24084	131	power	16733	181	created	14236
32	digital	39818	82	money	24028	132	allow	16588	182	website	14153
33	cost	37971	83	problem	23958	133	increase	16488	183	change	14133
34	financial	37912	84	distribution	23773	134	nodes	16225	184	node	14124
35	ecosystem	36681	85	set	23679	135	main	16207	185	peer	14041
36	customer	36035	86	protocol	23666	136	solutions	16195	186	securities	13997
37	whitepaper	35879	87	need	23448	137	various	16177	187	currencies	13982

(*continued on next page*)

(*continued*)

S.No.	word	freq	S.No.	word	freq	S.No.	word	freq	S.No.	word	freq
38	fee	35877	88	social	23266	138	form	16165	188	partners	13906
39	security	35685	89	account	22977	139	secure	16132	189	related	13872
40	trading	34244	90	distributed	22461	140	technical	16097	190	source	13828
41	access	34103	91	part	22434	141	operations	15989	191	provided	13759
42	price	33844	92	support	22177	142	existing	15975	192	marketplace	13712
43	community	33830	93	software	21878	143	energy	15960	193	version	13704
44	investment	33820	94	solution	21723	144	trade	15890	194	foundation	13689
45	industry	33243	95	mobile	21719	145	address	15886	195	demand	13688
46	funds	32650	96	different	21576	146	revenue	15859	196	group	13688
47	paper	32640	97	block	21409	147	initial	15830	197	businesses	13624
48	content	32631	98	offer	21404	148	billion	15738	198	developers	13607
49	marketing	32603	99	app	21379	149	buy	15644	199	regulatory	13566
50	wallet	32244	100	million	21197	150	advertising	15603	200	making	13513

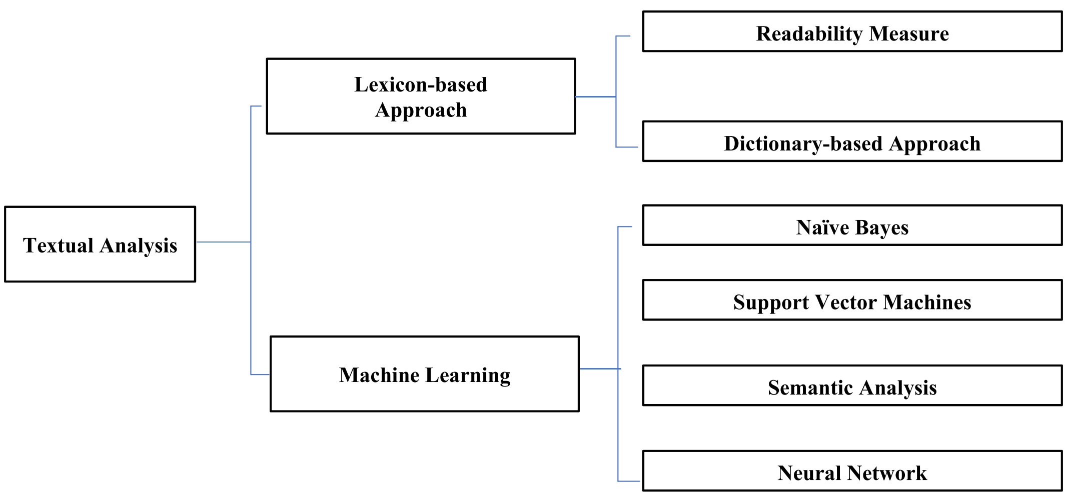
Appendix A.6.

Readability	Formula
Flesch	206.835 - (1.015 * number of words/number of sentences) - (84.6/number of syllables/number of words)
Flesch–Kincaid	0.39 * (number of words/number of sentences) + 11.8 * (number of syllables / number of words) – 15.59
COL	0.0588 * (Average number of letters per 100 words) – 0.296 * (Average number of sentences per 100 words) – 15.8
RIX	(Number of words with 7 characters or more) / (number of sentences)
FOG	((Average number of words per sentence) + (number of words of 3 syllables or more)) * 0.4
ARI	4.71 * (number of characters / number of words) + 0.5 * (number of words / number of sentences) - 21.43
SMOG	1.043 * sqrt (30 * number of words with more than two syllables / number of sentences) + 3.1291

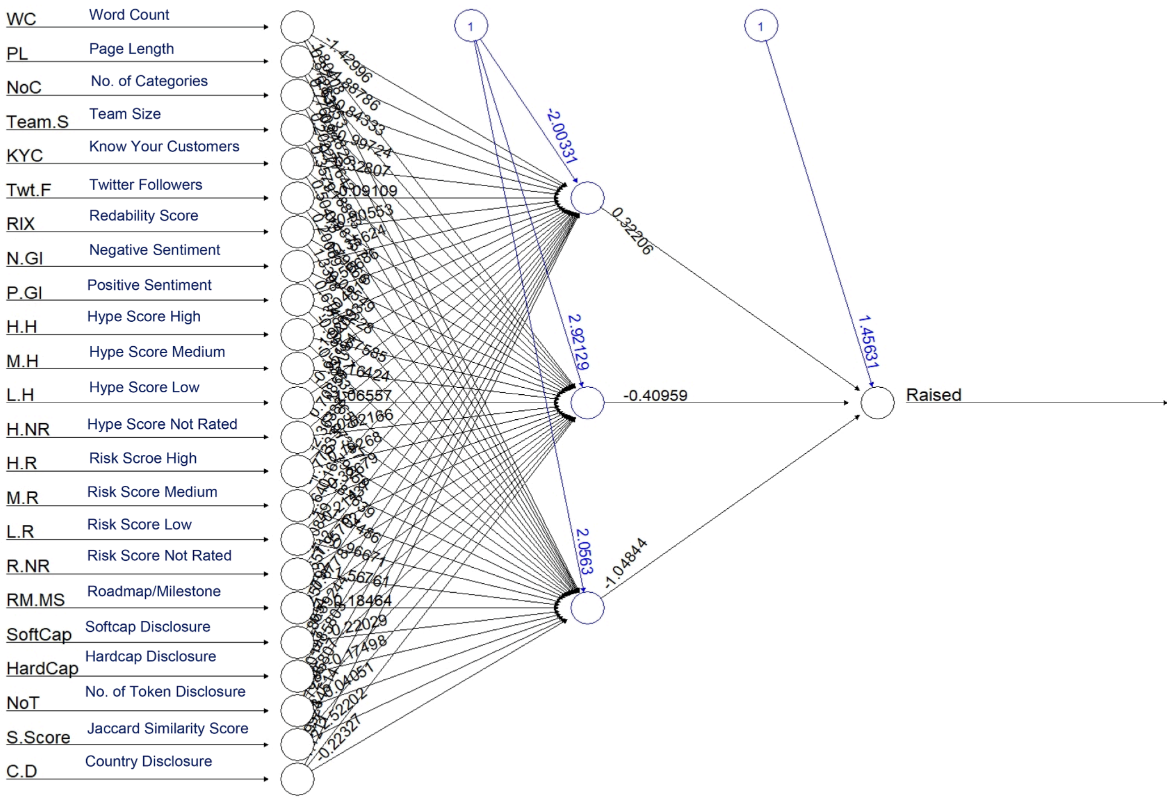
Appendix A.7. Flesh/Flesh-Kincaid Readability Measurement

Score	Notes
90-100	very easy to read, easily understood by an average 11-year-old student
80-90	easy to read
70-80	fairly easy to read
60-70	easily understood by 13- to 15-year-old students
50-60	fairly difficult to read
30-50	difficult to read, best understood by college graduates
0-30	very difficult to read, best understood by university graduates

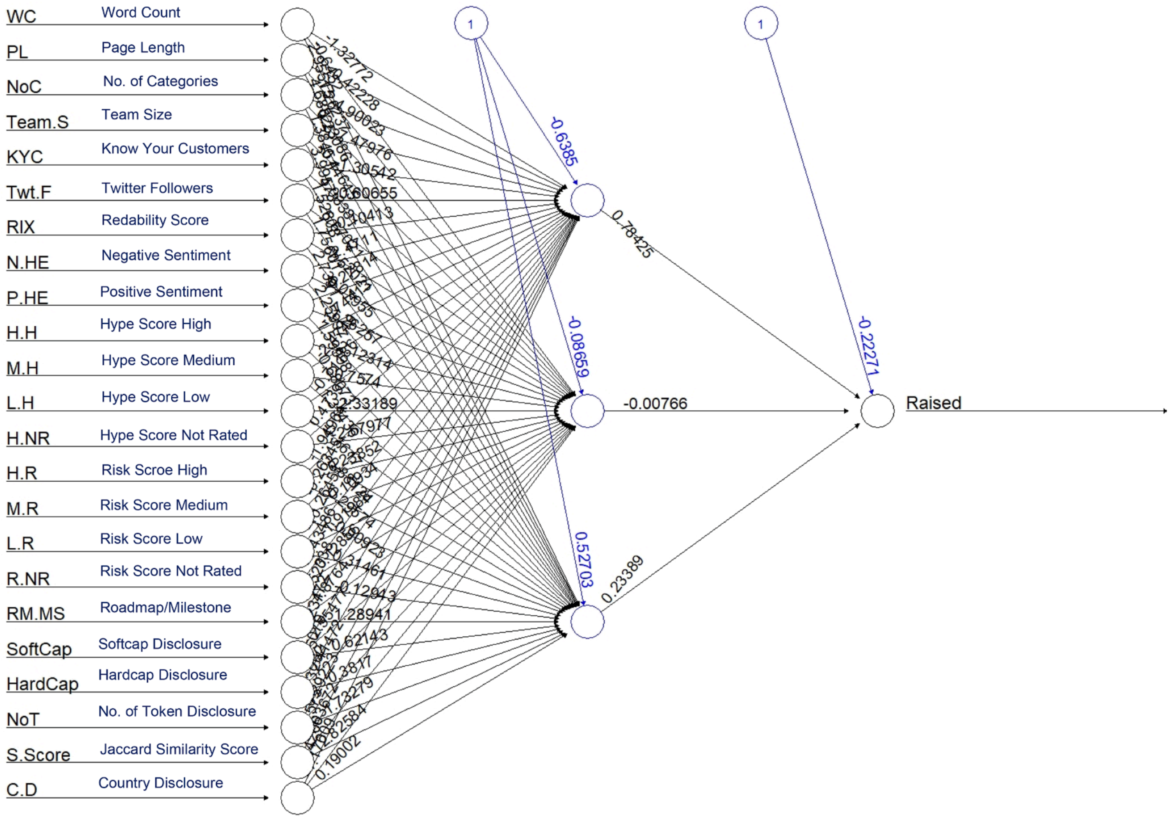
Appendix A.**8.** General Method for Textual Analysis (source: Guo et al., 2016)



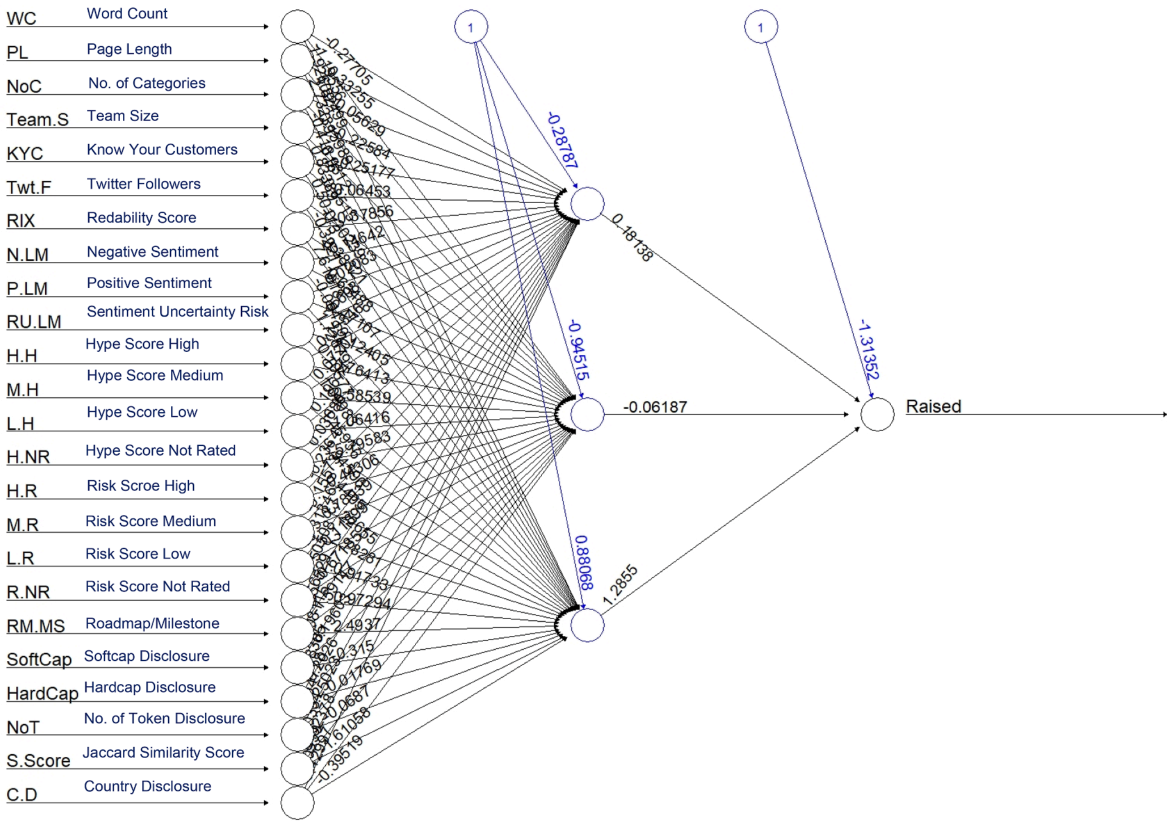
Appendix A.**9.** ANN Incorporating Harvard GI Sentiment Dictionary (*75% training and 25% test data with 3 hidden layers, RIX*)



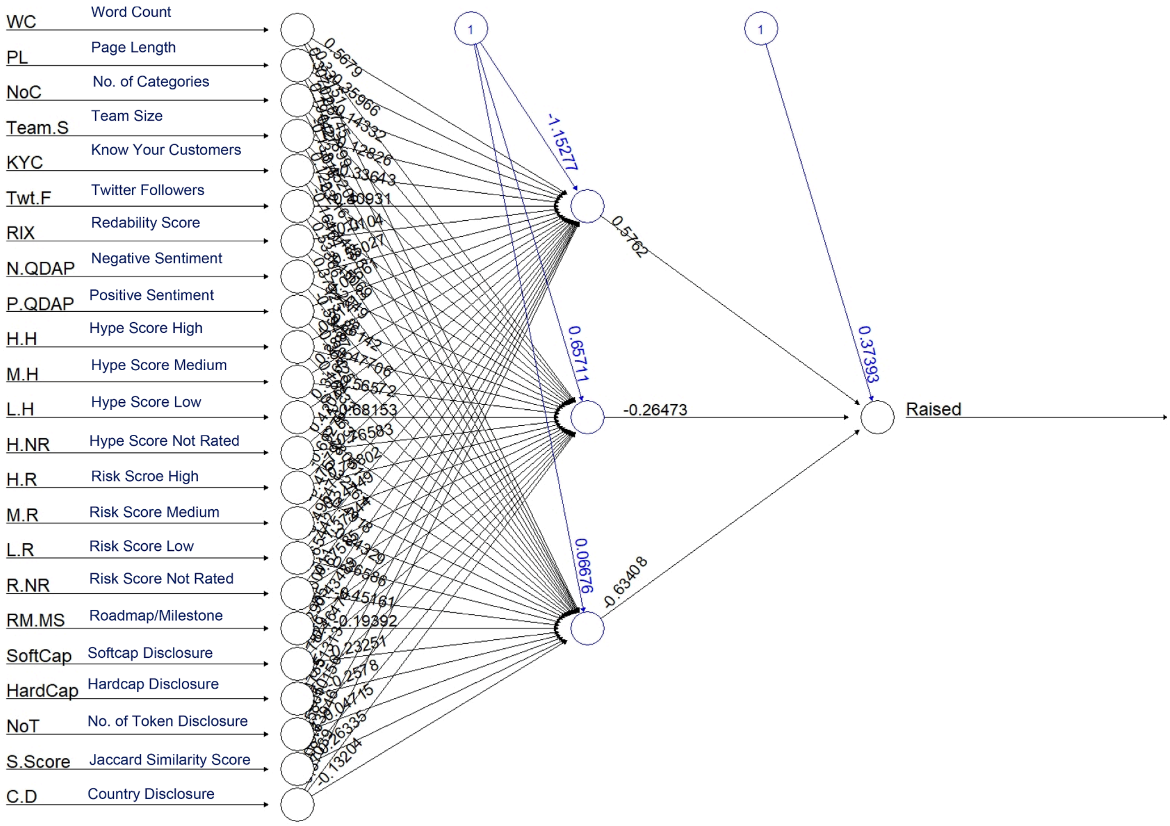
Appendix A.10. ANN Incorporating Henry's Finance Dictionary (*75% training and 25% test data with 3 hidden layers, RIX*)



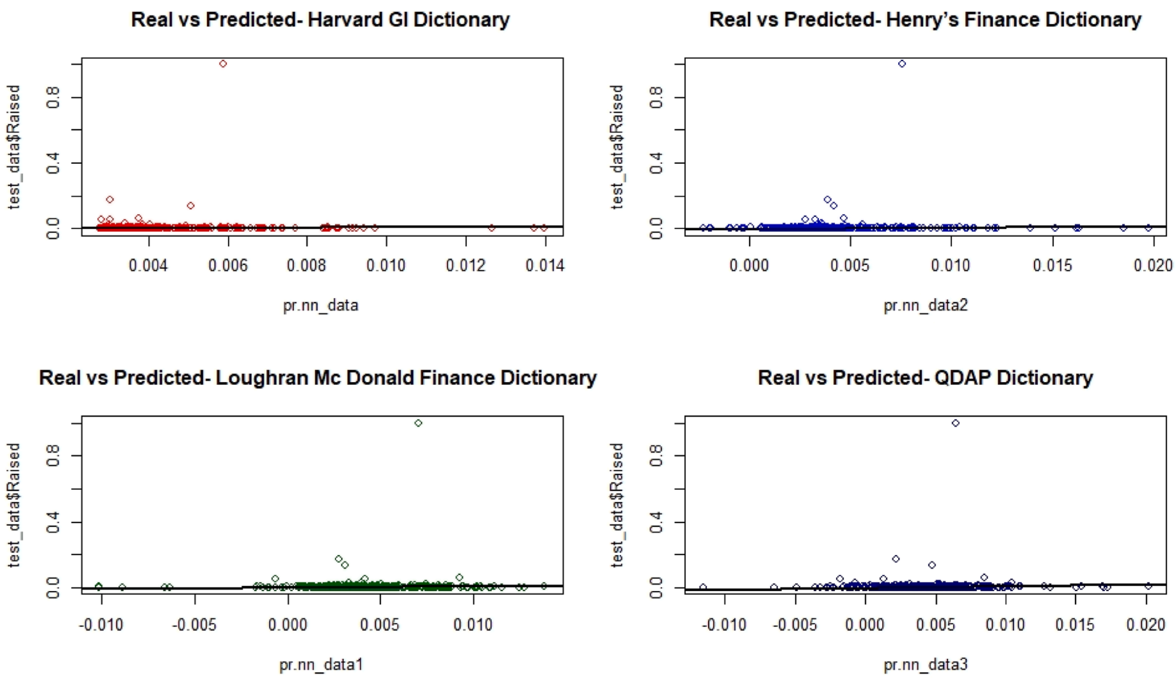
Appendix A.11. ANN Incorporating Loughran McDonald Finance Dictionary (*75% training and 25% test data with 3 hidden layers, RIX*)



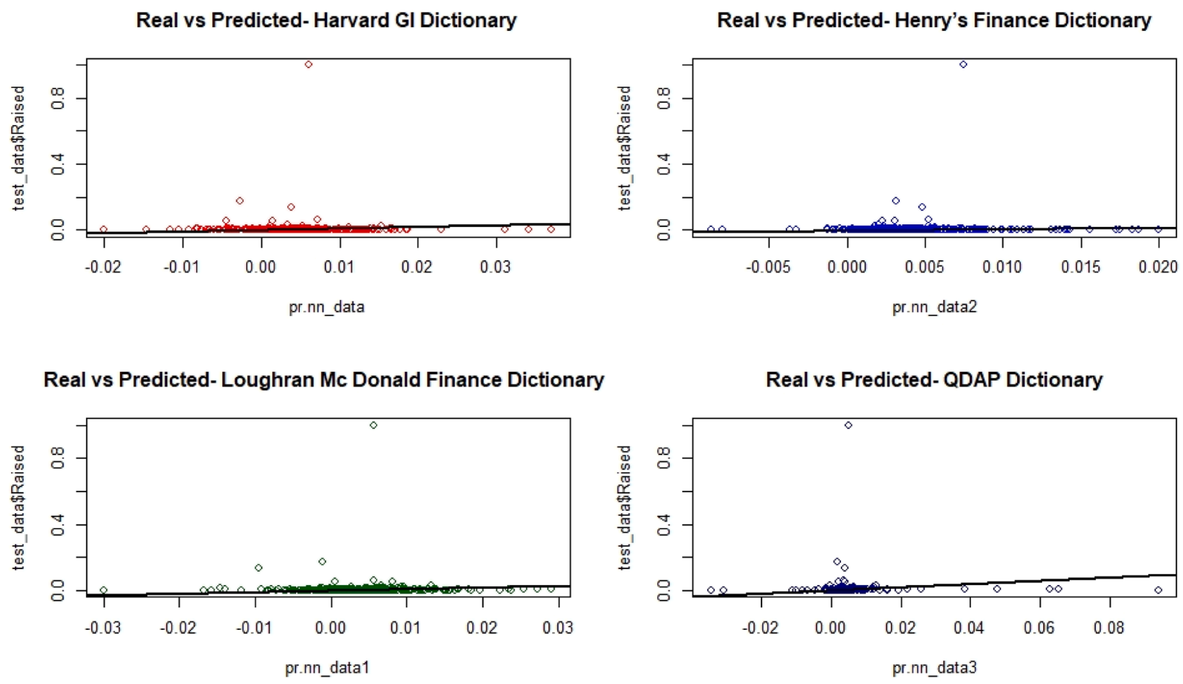
Appendix A.12. ANN Incorporating QDAP Sentiment Dictionary (*75% training and 25% test data with 3 hidden layers, RIX*)



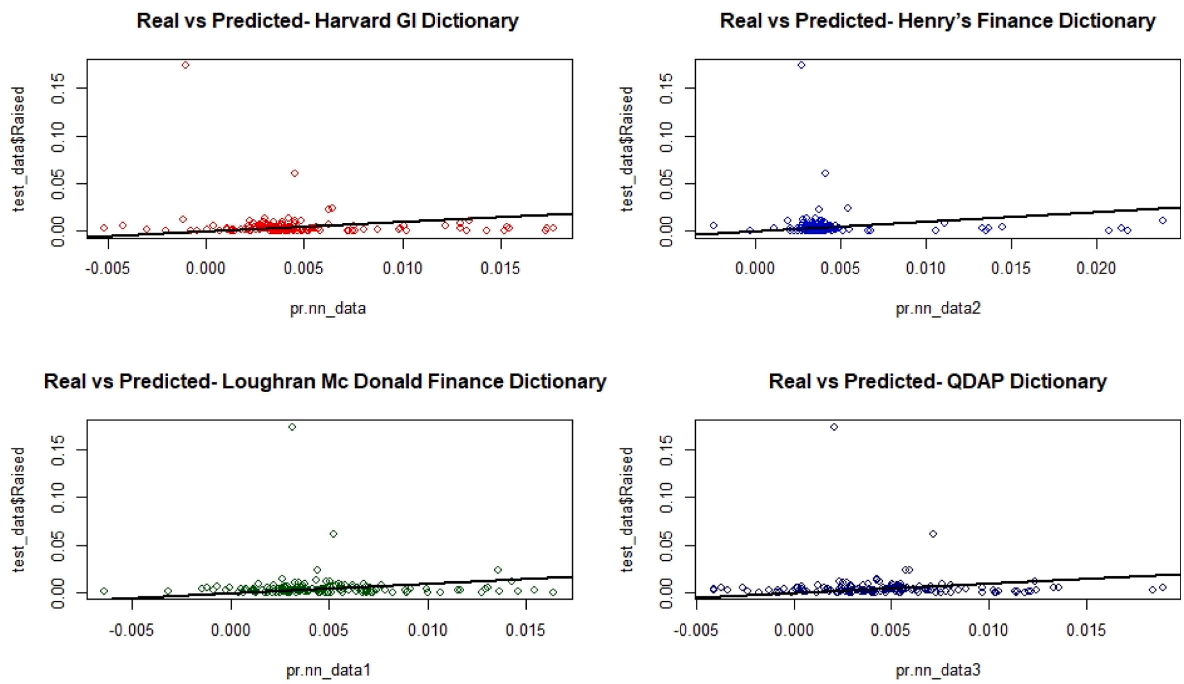
Appendix A.**13. ANN Model Fitting on Four Different Sentiment Dictionaries (***75% training and 25% test data with 3 hidden layers, and SMOG readability score***)**



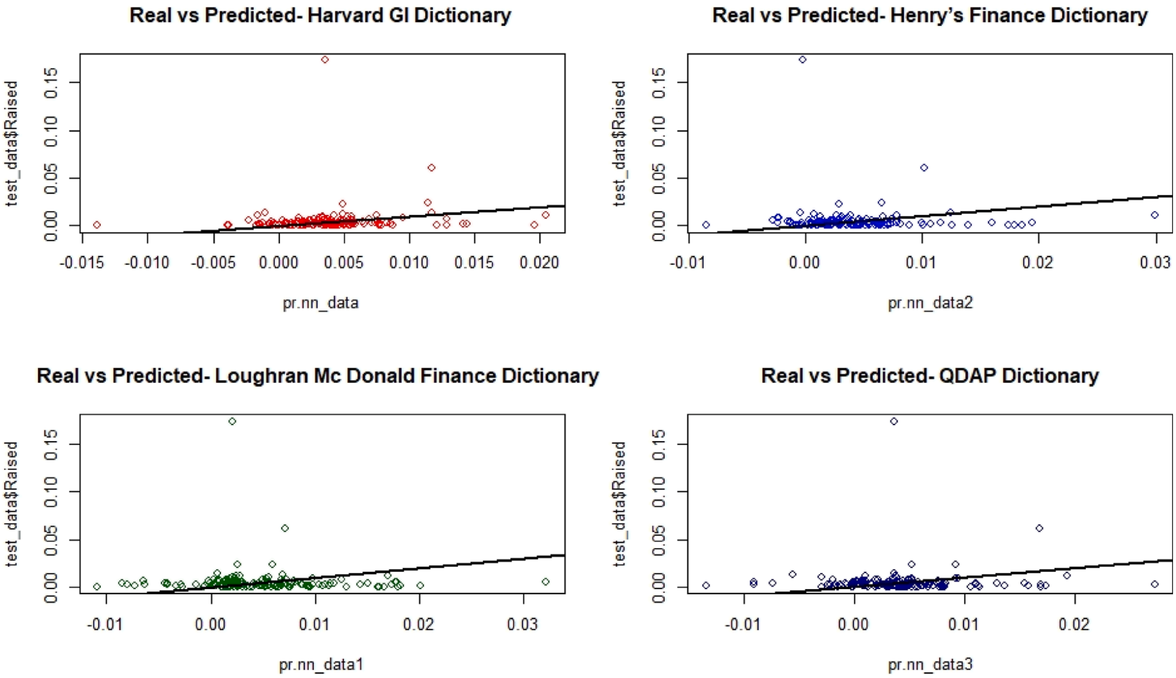
Appendix A.14**. ANN Model Fitting on Four Different Sentiment Dictionaries (**75% training and 25% test data with 10 hidden layers and 5 neurons, and SMOG readability score)



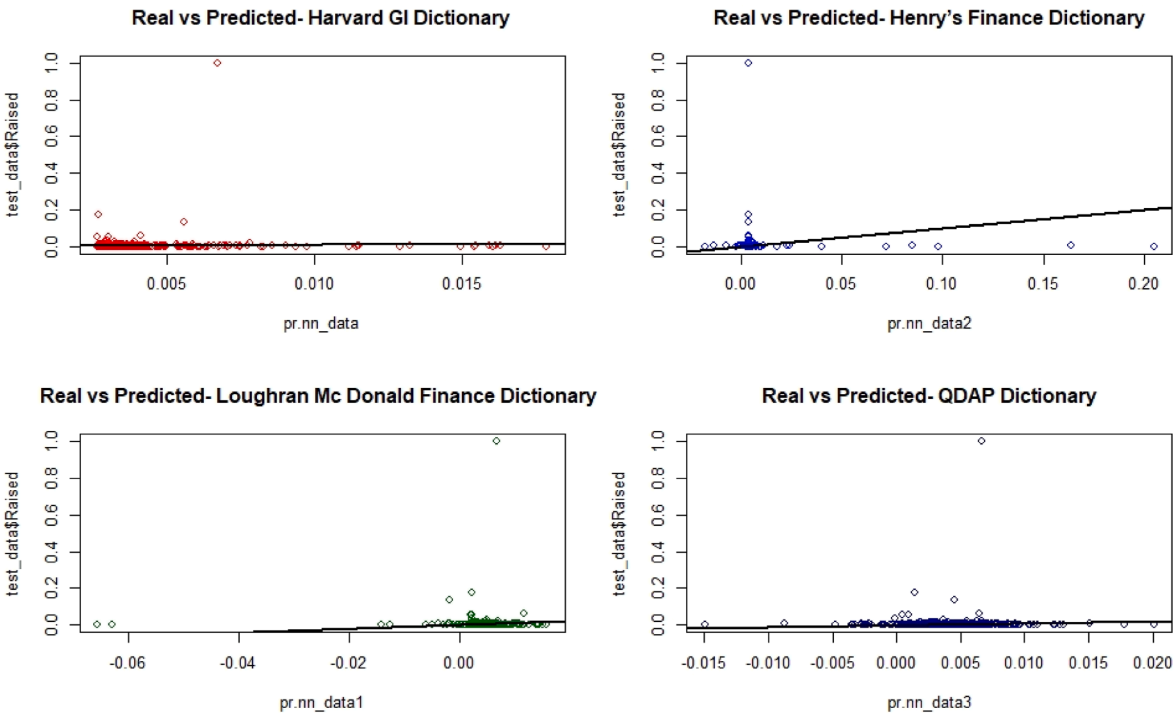
Appendix A.15**. ANN Model Fitting on Four Different Sentiment Dictionaries (**90% training and 10% test data with 3 hidden layers, and SMOG readability score**)**



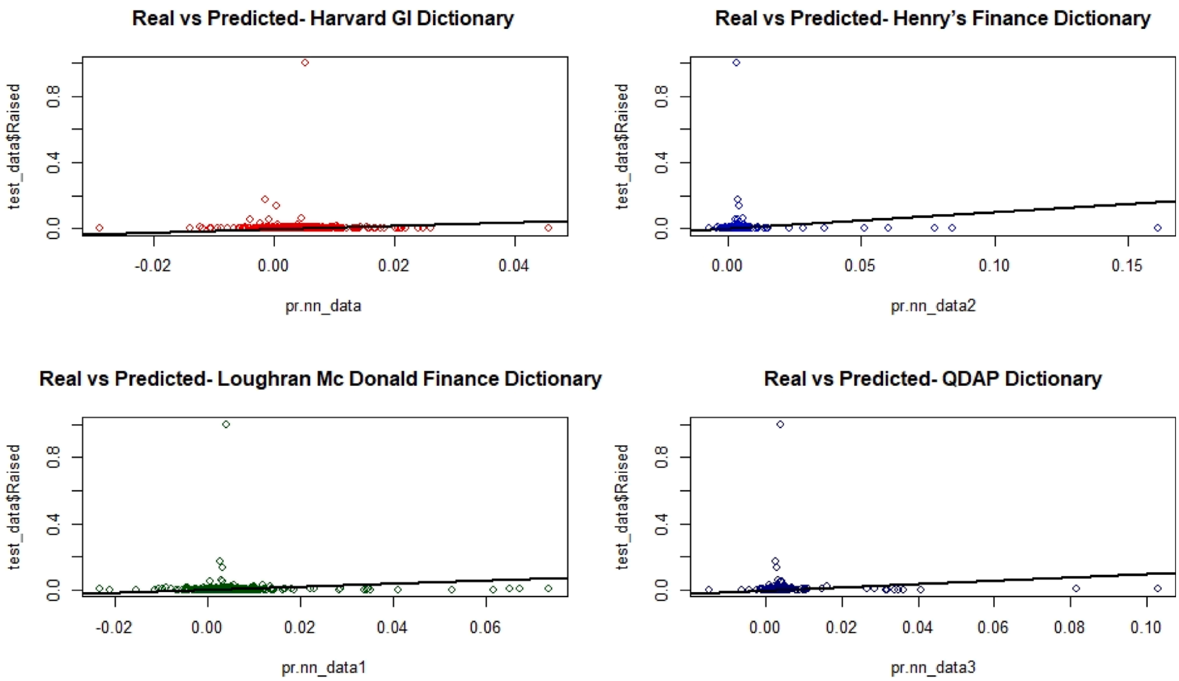
Appendix A.**16. ANN Model Fitting on Four Different Sentiment Dictionaries (**90% training and 10% test data with 10 hidden layers and 5 neurons, and SMOG readability score**)**



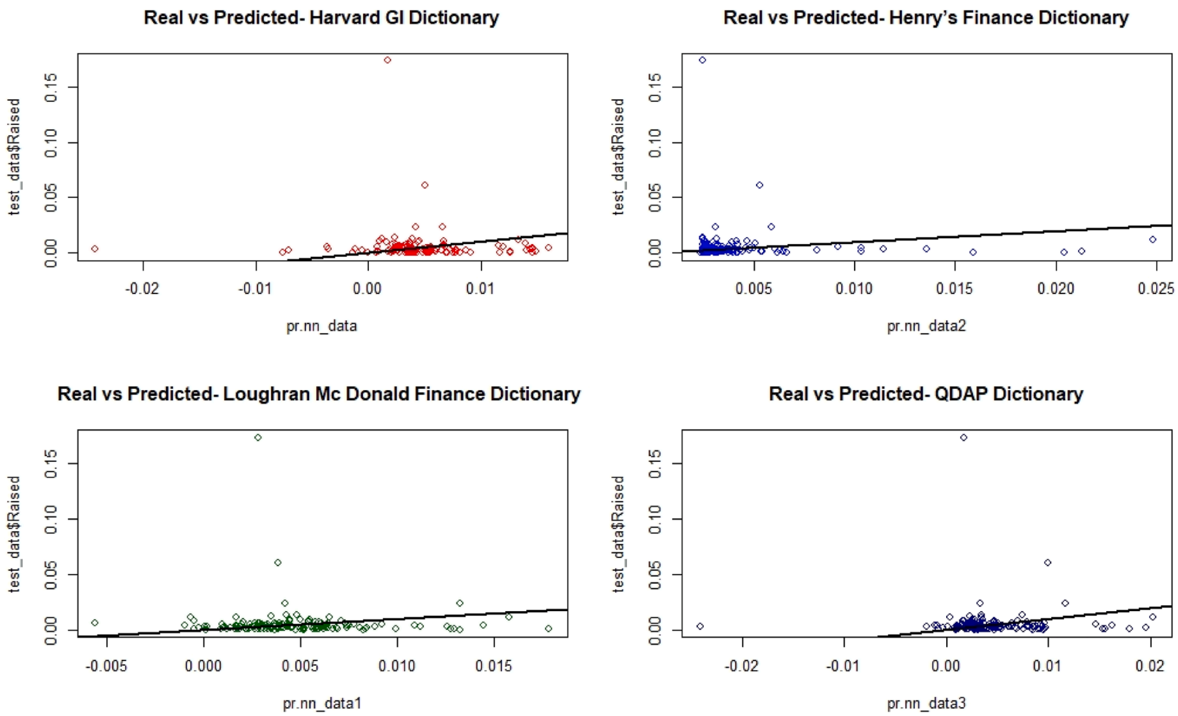
Appendix A.17. ANN Model Fitting on Four Different Sentiment Dictionaries (75% training and 25% test data with 3 hidden layers, and FOG readability score**)**



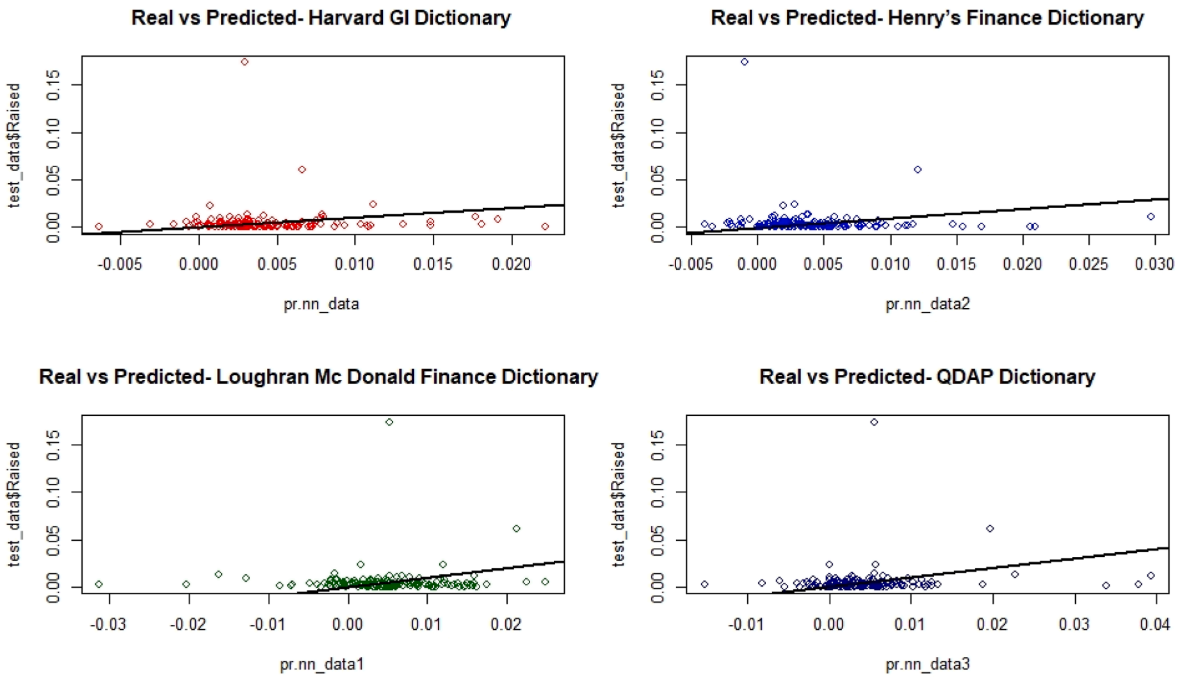
Appendix A.18**. ANN Model Fitting on Four Different Sentiment Dictionaries (**75% training and 25% test data with 10 hidden layers and 5 neurons, and FOG readability score**)**



Appendix A.19. ANN Model Fitting on Four Different Sentiment Dictionaries (90% training and 10% test data with 3 hidden layers, and FOG readability score**)**



Appendix A.20**. ANN Model Fitting on Four Different Sentiment Dictionaries (**90% training and 10% test data with 10 hidden layers and 5 neurons, and FOG readability score**)**



References

Adhami, S., Giudici, G., Martinazzi, S., 2018. Why do businesses go crypto? An empirical analysis of initial coin offerings. Journal of Economics and Business 100, 64–75.

Alessia, D., Ferri, F., Grifoni, P., Guzzo, T., 2015. Approaches, tools and applications for sentiment analysis implementation. International Journal of Computer Applications 125 (3).

Altarriba, J., 2006. Cognitive approaches to the study of emotion-laden and emotion words in monolingual and bilingual memory. Bilingual Education and Bilingualism 56, 232.

Amsden, R. and Schweizer, D., 2018. Are Blockchain Crowdsales the New'Gold Rush'? Success Determinants of Initial Coin Offerings. *SSRN (April 16, 2018)*.

Anooshian, L.J., Hertel, P.T., 1994. Emotionality in free recall: Language specificity in bilingual memory. Cognition & Emotion 8 (6), 503–514.

Bajo, E., Raimondo, C., 2017. Media sentiment and IPO underpricing. Journal of Corporate Finance 46, 139–153.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. The journal of Finance 61 (4), 1645–1680.

Baker, M., Wurgler, J., 2007. Investor sentiment in the stock market. Journal of economic perspectives 21 (2), 129–152.

Baker, S.R., Bloom, N., Davis, S.J., 2016. Measuring economic policy uncertainty. The quarterly journal of economics 131 (4), 1593–1636.

Benedetti, H., Kostovetsky, L., 2021. Digital tulips? Returns to investors in initial coin offerings. Journal of Corporate Finance 66, 101786.

Boreiko, D., Risteski, D., 2020. Serial and large investors in initial coin offerings. Small Business Economics 1–19.

Catalini, C. and Gans, J.S., 2018. Initial coin offerings and the value of crypto tokens. *National Bureau of Economic Research*, (No. w24418).

Chen, P., Lin, J., Chen, B., Lu, C., Guo, T., 2015. Processing emotional words in two languages with one brain: ERP and fMRI evidence from Chinese-English bilinguals. Cortex 71, 34–48.

Chiu, T., Chiu, V., Wang, T.D., Wang, Y., 2022. Using Textual Analysis to Detect Initial Coin Offering Frauds. Journal of Forensic Accounting Research.

Chod, J., Lyandres, E., 2021. A theory of icos: Diversification, agency, and information asymmetry. Management Science 67, 5969–6627.

Domingo, R.S., Piñeiro-Chousa, J., López-Cabarcos, M.Á., 2020. What factors drive returns on initial coin offerings? Technological Forecasting and Social Change 153, 119915.

Drobetz, W., Momtaz, P.P., Schröder, H., 2019. Investor sentiment and initial coin offerings. The Journal of Alternative Investments 21 (4), 41–55.

Fisch, C., 2019. Initial coin offerings (ICOs) to finance new ventures. Journal of Business Venturing 34 (1), 1–22.

Giudici, G., Moncayo, G.G., Martinazzi, S., 2020. The role of advisors' centrality in the success of Initial Coin Offerings. Journal of Economics and Business 112, 105932.

Grobys, K., Sapkota, N., 2020. Predicting cryptocurrency defaults. Applied Economics 52 (46), 5060–5076.

Grobys, K., King, T., Sapkota, N., 2022. A Fractal View on Losses Attributable to Scams in the Market for Initial Coin Offerings. Journal of Risk and Financial Management 15 (12), 579.

Guldiken, O., Tupper, C., Nair, A., Yu, H., 2017. The impact of media coverage on IPO stock performance. Journal of Business Research 72, 24–32.

Guo, L., Shi, F., Tu, J., 2016. Textual analysis and machine leaning: Crack unstructured data in finance and accounting. The Journal of Finance and Data Science 2 (3), 153–170.

He, W., Wu, H., Yan, G., Akula, V., Shen, J., 2015. A novel social media competitive analytics framework with sentiment benchmarks. Information & Management 52 (7), 801–812.

Henry, E., 2008. Are investors influenced by how earnings press releases are written?. *The Journal of Business Communication (1973)*, *45*(4), pp.363-407.

Hornuf, L., Kück, T., Schwienbacher, A., 2021. Initial coin offerings, information disclosure, and fraud. Small Business Economics 1–19.

Howell, S.T., Niessner, M., Yermack, D., 2020. Initial coin offerings: Financing growth with cryptocurrency token sales. The Review of Financial Studies 33 (9), 3925–3974.

Huang, D., Jiang, F., Tu, J., Zhou, G., 2015. Investor sentiment aligned: A powerful predictor of stock returns. The Review of Financial Studies 28 (3), 791–837.

Huang, W., Meoli, M., Vismara, S., 2020. The geography of initial coin offerings. Small Business Economics 55 (1), 77–102.

Jovanovic, B., Rousseau, P.L., 2001. Why wait? A century of life before IPO. American Economic Review 91 (2), 336–341.
Kaal, W.A., 2018. Initial Coin Offerings: The top 25 jurisdictions and their comparative regulatory responses (as of May 2018). Stan. J. Blockchain L. & Pol'y 1, 41.
Kissler, J., Herbert, C., Peyk, P., Junghofer, M., 2007. Buzzwords: early cortical responses to emotional words during reading. Psychological Science 18 (6), 475–480.
Li, J., Mann, W., 2018. Initial coin offerings and platform building. SSRN Electronic Journal.
Li, X., Xie, H., Chen, L., Wang, J., Deng, X., 2014. News impact on stock price return via sentiment analysis. Knowledge-Based Systems 69, 14–23.
Loughran, T., McDonald, B., 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of finance 66 (1), 35–65.
Loughran, T., McDonald, B., 2013. IPO first-day returns, offer price revisions, volatility, and form S-1 language. Journal of Financial Economics 109 (2), 307–326.
Lyandres, E., Palazzo, B. and Rabetti, D., 2019. Do tokens behave like securities? An anatomy of initial coin offerings. *SSRN Electronic Journal*.
Lyandres, E., Palazzo, B., Rabetti, D., 2022. Initial coin offering (ico) success and post-ico performance. Management Science.
Meyer, A., Ante, L., 2020. Effects of initial coin offering characteristics on cross-listing returns. Digital Finance 2 (3), 259–283.
Mohammad, S.M., Turney, P.D., 2013. Nrc emotion lexicon. National Research Council, Canada, p. 2.
Momtaz, P.P., 2020. Initial coin offerings. Plos one 15 (5), e0233018.
Muhammad, A., Wiratunga, N., Lothian, R., 2016. Contextual sentiment analysis for social media genres. Knowledge-based systems 108, 92–101.
Pavlenko, A., 2008. Emotion and emotion-laden words in the bilingual lexicon. Bilingualism: Language and cognition 11 (2), 147–164.
Pröllochs, N., Feuerriegel, S. and Neumann, D., 2015, May. Generating Domain-Specific Dictionaries using Bayesian Learning. In *ECIS*.
Qadan, M., 2019. Risk appetite, idiosyncratic volatility and expected returns. International Review of Financial Analysis 65, 101372.
Roosenboom, P., van der Kolk, T., de Jong, A., 2020. What determines success in initial coin offerings? Venture Capital 22 (2), 161–183.
Sapkota, N., 2022. News-based sentiment and bitcoin volatility. International Review of Financial Analysis 82, 102183.
Tsukioka, Y., Yanagi, J., Takada, T., 2018. Investor sentiment extracted from internet stock message boards and IPO puzzles. International Review of Economics & Finance 56, 205–217.
Wu, C., Zhang, J., Yuan, Z., 2021. Exploring Affective Priming Effect of Emotion-Label Words and Emotion-Laden Words: An Event-Related Potential Study. Brain Sciences 11 (5), 553.
Yap, M.J., Seow, C.S., 2014. The influence of emotion on lexical processing: Insights from RT distributional analysis. Psychonomic bulletin & review 21 (2), 526–533.
Yekrangi, M., Abdolvand, N., 2021. Financial markets sentiment analysis: Developing a specialized Lexicon. Journal of Intelligent Information Systems 57 (1), 127–146.
Young, L., Soroka, S., 2012. Affective news: The automated coding of sentiment in political texts. Political Communication 29 (2), 205–231.
Yu, G.N., 2019. Factors of success of Initial coin offering. Empirical evidence from 2016–2019. Финансы: теория и практика 23 (5).
Zetzsche, D.A., Buckley, R.P., Arner, D.W., Fohr, L., 2019. The ICO gold rush: It's a scam, it's a bubble, it's a super challenge for regulators. Harv. Int'l LJ 60, 267.
Zhang, D., He, W., Wang, T., Luo, W., Zhu, X., Gu, R., Li, H., Luo, Y.J., 2014. Three stages of emotional word processing: an ERP study with rapid serial visual presentation. Social cognitive and affective neuroscience 9 (12), 1897–1903.
Zhang, S., Aerts, W., Lu, L., Pan, H., 2019. Readability of token whitepaper and ICO first-day return. Economics Letters 180, 58–61.
Zhang, S., Aerts, W., Zhang, D., Chen, Z., 2022. Positive tone and initial coin offering. Accounting & Finance 62 (2), 2237–2266.
Zhao, W., Chen, L., Zhou, C., Luo, W., 2018. Neural correlates of emotion processing in word detection task. Frontiers in psychology 9, 832.