“Get a higher return on your savings!”: Comparing adverts for cryptocurrency investment scams across platforms

Gilberto Atondo Siu
Computer Laboratory
University of Cambridge
Cambridge, United Kingdom
jga33@cam.ac.uk

Alice Hutchings
Computer Laboratory
University of Cambridge
Cambridge, United Kingdom
alice.hutchings@cl.cam.ac.uk

***Abstract*—This work compares machine learning methods using supervised, semi-supervised and unsupervised learning, to classify advertisements for cryptocurrency related investment scams found in the online forum Bitcointalk, and the social media platform Reddit. We extract more than 24.2 million posts from Bitcointalk and use Reddit’s API to collect 2,108 submissions. We train and compare several multi-class text classification approaches and use the models with highest accuracy and F-measure to identify cryptocurrency investment scam advertisements found on both platforms. We discover around five percent of all posts collected on both sites are potential scams. We then use another text classifier to identify the scam actors involved in these investment scam advertisements. We also discover the lures used within these fraudulent adverts and find the main differences in luring techniques used between Bitcointalk and Reddit. We identify that the most prevalent lure type uses the financial principle, followed by the distraction principle in Bitcointalk, and by the authority principle in Reddit. Finally, we use subreddits as communities’ proxies and compare scam advertisements within them to identify whether pensioners are being specifically targeted by cryptocurrency scam adverts. Our results show that retirement subreddits do not contain a higher number of cryptocurrency investment scam adverts compared to other investment focused subreddits.**

***Index Terms*—cryptocurrency, cybercrime, investment scams, machine learning**

1. Introduction

Cryptocurrencies have become increasingly popular over the past few years. It is estimated that the number of cryptocurrency users worldwide has increased by more than 80 fold from 5 million in 2016 to 402 million in 2022 [49]. The cryptocurrency ecosystem has also enabled the creation of decentralised finance (DeFi) which aims to replicate many services of ‘traditional’ finance in an autonomous and permissionless manner. Many investors have flocked into this space. According to a report from the OECD, the total value of the DeFi market increased from $1.9 billion in July 2020 to $100 billion in November 2021 [39].

Some users [7] think of cryptocurrencies as a long term investment asset, on occasions compared to gold [17], which allows them to increase their portfolio diversification [18]. Other users [14] see them as a mean to make a quick profit by taking advantage of high volatility in cryptocurrency prices or participating in pump and dump schemes [21]. Fraudsters have taken advantage of this appetite for risk and reward offered by cryptocurrencies. According to the FBI, losses related to cryptocurrency investment crime increased by 183% from $907 million in 2021 to $2.57 billion in 2022 [20].

In addition, even though some regulatory frameworks have been proposed [19], [32], absence of globally coordinated regulation and transparency in DeFi and the cryptocurrency ecosystem continues to persist. This lack of oversight includes advertisement of cryptocurrency investment in online platforms. Criminals leverage the popularity of these sites and can use social engineering techniques to promote fake investment schemes and defraud investors within them.

This work focuses on cryptocurrency investment scam advertisements. We define these as conversations found on online forums or social media, that promote cryptocurrency-based schemes offering an extremely high rate of return within a very short investment period (1-2% daily). These schemes can sometimes work as ‘pyramid’ schemes, where earnings of early investors originate from investments of late investors, until the program collapses. On some occasions, these schemes also promise a guaranteed investment return due to the use of obscure and complex methods [45].

Researchers have investigated these types of schemes from different angles. For example, some researchers have focused in analysing success factors of programs [36], [54] and others have investigated their profitability [16], [37]. However, there is less extensive work around the luring techniques used in the advertisement of these schemes. In particular and to the best of our knowledge, there is no research done on cryptocurrency scams advertisements targeted specifically at pensioners.

Atondo Siu et al. [3] provide a longitudinal analysis of investment scam lures found on Bitcointalk, the longest-running and most popular cryptocurrency-themed online forum, from 2010 to 2022. Statistical models are used to categorise more than 281k threads, with around six percent of threads predicted to be investment scam advertisements. The types of actors and lures being used to facilitate the scams are identified, with the promise of financial gain being the most commonly used tactic by criminals to attract victims. In this paper, we build on that work.

Identification of cryptocurrency investment scams adverts with a high level of accuracy is important. Therefore, one of our objectives is to find a method that is the best predictor of cryptocurrency investment scam advertisements. For this purpose, we compare the performance of several machine learning methods, using supervised, semi-supervised, and unsupervised learning. We start with a thread type classifier to identify cryptocurrency investment scam adverts and scam-related comments. We compare the performance of the following models: XGBooost [42], Long-short Term Memory (LSTM) [23], Convolutional Neural Networks (CNN) [29] with LSTM (CNN-LSTM) and Zero-Shot learning [28]. We also test several versions of LSTM with Global Vector for word representation (GloVe) [41] using ten different pre-trained word vectors from Stanford-NLP. We select the model with the highest accuracy and F-measure and implement active learning to check whether the performance can be improved.

We then analyse the prevalence of adverts for cryptocurrency investment scams in two platforms. We use Bitcointalk because is the longest running cryptocurrency-focused online forum. We leverage the work done by Atondo Siu et al. [3] who collected more than 17.8 million posts from this site. We expand this and our increased dataset contains, as of March 2023, more than 24.2 million posts from 535,300 threads, from November 2009 until February 2023. We also use Reddit because it has become an important platform for investors in the last few years. One of the reasons behind this trend is because this social media channel allows users to gather in communities or subreddits specifically dedicated to investing and other activities. We use Reddit's API to collect 2,108 Reddit posts.

After identifying all cryptocurrency investment scam adverts and scam-related comments on both data sources, we predict the post author (scam owner, shill, participant, victim, reporter, or commenter). We use an XGBoost model and implement active learning [47] to test whether we can increase the accuracy and F-measure. We then investigate the investment scam lures (using Stajano & Wilson's typology [48]) used within the scam adverts by using multiple logistic regression classifiers. We finish our research by finding the differences in lures used in scam advertisements on Bitcointalk and Reddit.

The final objective of this work is to understand whether pensioners are targeted more heavily with investment scam advertisements than other communities for investors. Pensioners are being increasingly targeted by cybercriminals. Researchers [9], [33], [38] have found that elderly people are targeted more frequently by scammers, compared to the rest of the population, for several reasons. Some of these include a higher number of physical and mental health issues, seclusion and cognitive problems. Pensioners are also perceived to have higher levels of accumulated wealth along with a lower level of cybersecurity understanding and protection [9].

Regulators and other governmental agencies have also raised concerns about the risks that pensioners can face due to their perceived accumulated lifetime savings [50]. In 2021, the FBI found that more than 92,000 people over the age of 60, in the United States, reported fraud-related losses of more than $1.7 billion to the Internet Crime Compliant Center. This specific age group had the largest number of victims and reported the biggest losses out of all groups. In particular, the losses related to investment scams accounted for $239 million, many of which involved cryptocurrencies [25].

Based on the above, it is important to detect these schemes proactively so elderly users can be warned. In addition, knowing the tactics used by fraudsters to lure victims into these programs can help pensioners avoid losing their lifetime savings. We therefore use subreddits as communities' proxies to compare scam adverts targeted at investors in general with those targeted at pensioners. We identify four active subreddits with the largest number of subscribers focused on investing (r/investing, r/investment and r/InvestmentClub) and retirement (r/retirement) and compare whether there is a difference in the number of scam adverts found within them.

The goal of our research is to address the following questions:

1) How do statistical models such as XGBoost compare to other machine learning methods for the prediction of investment scam advertisements in Bitcointalk and Reddit?
2) How prevalent are cryptocurrency investments scams advertisements in both platforms?
3) What are the differences (if any) in the number of investment scam adverts found within retirement-related subreddits in comparison with other subreddits focused in investment?

Our work is organised as follows. In §2, we review related work. Our methods are shown in §3, including details of our data collection, ethical considerations and our classifiers. We then present the classification results and discuss them in §4 before providing our conclusions in §5.

2. Related Work

2.1. Cryptocurrency-related scams

Extensive work has been done to investigate cryptocurrency scams. Badawi and Jourdan [4] provide a systematic review of publications focused on different cryptocurrency-related fraudulent schemes. They list the public datasets used by some of the papers, and the risks and proposed solutions mentioned in these publications. Trozze et al. [53] also perform a methodical review of cryptocurrency-related crime. They argue that there is inconsistency of definitions around crime types found in their analysis. They find that the majority of the literature they investigated is focused on Ponzi schemes, High Yield Investment Programs (HYIPs) and Initial Coin Offerings (ICOs).

Some of the earliest research on HYIPs was conducted by Moore et al. [36] who analyse tools used to promote these schemes and their longevity factors. Subsequently, Drew and Moore [16] evaluate links between these schemes' websites by employing clustering methods. Neisius and Clayton [37] also evaluate the profitability of HYIPs. Toyoda et al. [52] use Bitcoin addresses transactions linked to HYIP owners to train a classifier and predict whether a Bitcoin address belongs to specific HYIP administrators.

Vasek and Moore [54] classify Bitcoin-related scams into Ponzi schemes and three other types of scams. They analyse schemes' success factors and identify that a few victims are the source of large scam amounts. Later on, Vasek and Moore [55] link success characteristics to the schemes' life span and find that actors' reputation and high shill intervention are linked to schemes' with longer life span.

Chen et al. [12], [13] and Bartoletti et al. [6] investigate Ponzi schemes based on other cryptocurrencies such as Ethereum. Badawi et al. [5] investigate a type of Ponzi scheme where websites pretend to create new Bitcoins. Boshmaf et al. [8] analyse Bitcoin address transactions, found within conversations on Bitcointalk, that are linked to a specific Ponzi scheme. Ibba et al. [24] analyse Ethereum smart contracts and use statistical methods to classify 'Ponzi scheme contracts'.

Other work analyses a variety of specific cryptocurrency-related crime. For example, Sapotka et al. [43] analyse fraudulent ICOs, Mazzorra et al. [34] and Agarwal et al. [2] focus on rug pulls, and Kshetri [27] investigate Non-Fungible-Tokens-related scams.

2.2. Social engineering techniques and studies of older online users

Xia et al. [57] analyse a number of cryptocurrency scams related to the COVID-19 pandemic. They find 195 fraudulent schemes and analyse the social engineering methods used in these programs and estimate monetary losses linked to the schemes' blockchain transactions. Weber et al. [56] also investigate the use of social engineering techniques within five cryptocurrency-related fraud cases. Mackenzie [31] performs an ethnographic study of cryptocurrency trading, arguing that legitimate programs and investment scams are indistinguishable.

Stajano and Wilson [48] provide a typology of lures used by online scammers. These lures include the authority, dishonesty, distraction, financial, herd, kindness, and time principles. This typology was applied to cryptocurrency investment scams by Atondo Siu et al. [3], providing a longitudinal analysis of scam lures found on Bitcointalk. While the financial principle was the most commonly used lure, there was a gradual rise in the use of the distraction principle. The kindness principle was used more frequently during the COVID-19 pandemic. Contrary to claims by Mackenzie [31], the time principle was not used as frequently as expected.

Martin and Rice [33] analyse testimonies from elderly online users and state that high net worth individulas are more susceptible to be targeted by online scammers. Nicholson et al. [38] investigate cybersecurity awareness of older internet users and discover that this group do not prioritise cybersecurity information when they use online resources. Burton et al. [9] provide a victimisation review of the risk factors that make elderly internet users more susceptible to financial cybercrime than other age groups. They state that mental health issues along with physical and cognitive impairment are among the causal factors that lead to victimisation of this group.

2.3. Automated text classification of social media posts

Zahrah et al. [59] use topic modelling and sentiment analysis to investigate online hate speech in Reddit and 4chan. They find that these platforms can sometimes be used to influence users to vote in a certain manner. Zhou et al. [60] also focuses on analysing hateful content but they use underground and extremist forums as their data source. They find that a classifier trained using multiple data sources does not always outperform classifiers that use a single data source. Ismail and Yusoff [26] focus on multi-class categorisation of gender violence and implement a hybrid LSTM-CNN model with GloVe word embeddings on Twitter posts.

Sarabadani et al. [44] provides a longitudinal study of COVID-19 symptoms documented by patients on Reddit posts. They use active learning and state that it is possible to use social media data to get a better understanding of disease evolution and the physical and psychological symptoms. Tokala et al. [51] compare several statistical models with deep learning methods, including LSTM, LSTM with Glove word embeddings and CNN-LSTM, to analyse Twitter data related to medication intake and mentions of drugs and dietary supplements. Other authors explore the use of active learning and deep learning models to deal with unbalanced datasets [30], [35].

3. Methods

Analysing and classifying text from online forums and social media can be challenging and many methods have been proposed to automatically classify this type of text ([15], [22], [26], [44], [51], [59], [60]). One of our objectives is to find the best method to predict cryptocurrency investment scam advertisements. We start by using the pre-processed annotated data and create three text classifiers: one for scam thread type, another one for scam actor type and one more for scam lure type.

We find and use the best scam prediction method and categorise all posts from both data sources. After identifying investment scam advertisements in Bitcointalk and Reddit, and the actors behind them, we investigate the lures used within these posts and find the differences in lures between both platforms.

Another objective of this work is to identify whether pensioners are targeted more frequently than other age groups. We use subreddits from Reddit as communities' proxies and we check for any differences between the number of advertisements found in subreddits for pensioners and those aimed at investors in general.

3.1. Data collection

We use two different platforms to source our data. These platforms represent a variety of online users interested in investment. We leverage the work done by Atondo Siu et al. [3] who collected more than 17.8 million posts from Bitcointalk, the first (and longest-running) online forum created by and for users interested in discussions about Bitcoin and subsequent cryptocurrencies. We expand this dataset and our increased dataset contains, as of

March 2023, more than 24.2 million posts from 535,300 threads, from November 2009 until February 2023. We remove all non-English written posts and use 531,356 threads for our prediction of cryptocurrency investment scam adverts. To collect the data, we leveraged the data collectors from the Cambridge Cybercrime Centre[1] [40].

Our second data source is Reddit, which is the tenth most popular social media platform globally [46] used for news gathering, topic discussion and rating content. The site is composed by 'subreddits' or communities which are boards created by users, dedicated to specific themes. We choose Reddit as our second data source as it has become an important platform for investors because it allows them to gather in communities, some of them specifically dedicated to investing.

One of our objectives is to understand whether pensioners are targeted more heavily with investment scam advertisements than other investor communities. We use subreddits as communities' proxies to identify scam adverts targeted at investors in general and pensioners. Even though this method has some limitations (for example r/retirement is not limited to pensioners and, vice versa, pensioners can be active in other non-retirement subreddits such as r/investing, etc.), it allows us to compare whether there is any difference in the number of scam adverts specifically targeted towards different communities. Therefore, we use Reddit's API to extract posts from the four most popular and active subreddits focused on investing (r/investing, r/investment and r/InvestmentClub) and retirement (r/retirement). The API limits the number of submissions that can be collected per request so we extracted all submissions posted between 1 January and 1 March 2023 from each subreddit, totalling 2,108 posts.

3.2. Data annotation

We used the annotated dataset created by Atondo Siu et al. [3] as the ground truth, which includes 4,218 posts from 2,630 threads on Bitcointalk based on the criteria listed below. In addition, we annotate 150 additional threads during the implementation of active learning described in subsubsection 3.4.3.

The annotation and classification criteria for thread type are:

1) Overt scam. The thread invites others to invest in a scheme explicitly recognised as a scam (Ponzi scheme, HYIP, etc.). The thread title usually has the name/details of the scheme.
2) Potential scam. The thread invites others to invest in a scheme promising investment returns that are unusually high and/or guaranteed but it does not make specific reference to a Ponzi scheme or a HYIP. We include in this category advertisements for ICOs, cryptocurrency exchanges, mining companies, raffles and gambling adverts only if they offer high rates of return. We do not consider cryptocurrency mixers as investment scams.
3) Scam comment. Relates to investment scams, but is not an invitation to invest. May include people asking for advice for setting up scam-related investments, sharing advice on how to spot a scam, reporting a fraudulent or fake investment scheme (known or unknown as a Ponzi scheme) etc.
4) Not investment scam related. The post content or the thread title is not related to investment scams.

This criteria allows us to categorise adverts as overt cryptocurrency scams when they blatantly promote fraudulent schemes. At the same time, it provides the flexibility to classify as potential cryptocurrency scams, any other scheme that offers an extremely high rate of return within a short period of time, without explicitly accepting to be a fake scheme.

All threads classified as overt or potential cryptocurrency investment scam adverts or scam comments are then annotated to identify the type of actor behind the post using the criteria below. We also used the annotated dataset created by Atondo Siu et al. [3] which includes 1,313 posts. Additionally, we annotate 200 additional posts during the implementation of active learning described in subsubsection 3.4.4.

1) Scam owner. The user invites others to invest in an overt or potential scam. This is usually the first user who starts a thread.
2) Scam shill. This refers to users that make comments (usually positive) that seem to legitimise a scheme.
3) Scam participant. The user responds to the scam owner knowing that they are participating in a fraudulent scheme.
4) Scam victim. The user claims to have been defrauded or lost from a previous investment.
5) Scam reporter. The user reports other users claiming that their posts/advertisements are scams. They do not need to have invested and lost themselves.
6) Scam commenter. The user discusses investment scams but does not fall into any of the above categories.

We also categorise all scam related threads by lure type using Stajano and Wilson's [48] typology. We define the scam lures as:

1) Authority principle. The scammer invokes authority such as by demonstrating technical knowledge (e.g. using encryption) or referring to trusted third parties (e.g. Companies House, CloudFlare) to convince users to do things that they would not do otherwise.
2) Dishonesty principle. The scammer encourages others to participate by making them aware that their profit comes from the losses of others.
3) Distraction principle. The scammer offers an investment opportunity and provides a lot of irrelevant details.
4) Financial principle. The scammer takes advantage of users' 'need and greed' to promise enticing options and convince users to make an investment.
5) Herd principle. The scammer refers to the popularity of the scheme to convince victims to not be left out of the investment rewards.

[1] https://www.cambridgecybercrime.uk

6) Kindness principle. The scammer relies on users' willingness to help in order to steal their money.
7) Time principle. The scammer puts time pressure on users so they make rushed and less reasonable choices.

3.3. Ethical considerations

This research was approved by the department's ethics committee at the University of Cambridge. The data used for this work was collected from an online forum and a social media platform, both of which are publicly available. The online forum provides privacy advice to users, indicating they are aware postings are public. The forum allows users to participate in it without submitting any personal information and it does not seem to have private information that has become publicly available inadvertently. The online forum's terms of service do not explicitly forbid scraping. Data from Reddit were collected using their API. This work focuses on understanding aggregate information and collective behaviour. We do not investigate specific individuals, or attempt to identify forum users. Therefore, this work falls outside the requirement of informed consent, under the British Society of Criminology's Statement of Ethics [1].

3.4. Text classification

3.4.1. Data Pre-processing. The annotated dataset was pre-processed by removing blank inputs, punctuation, stopwords and website links, We also converted all text to lower case, and tokenised it.

We used a ratio of 67/33 to split the input data for training and testing correspondingly. The majority of posts in Bitcointalk and Reddit are not investment scam related. The training sample was obtained randomly to reflect an accurate representation of the investigated platforms and therefore it is unbalanced. We oversampled the training data using SMOTE [11] to deal with the skewed data distribution.

3.4.2. Performance Measures. To compare the performance of all the methods used in our research, we used measures of accuracy, precision and recall. All of these scores range from 0, the worst possible score, to 1, the best. Accuracy is defined as the percentage of correct predictions.

$$\text{Accuracy} = \frac{Numberofcorrectpredictions}{Totalnumberofpredictions} \tag{1}$$

Precision is defined as the percentage of threads/posts that are correctly and positively categorised by the model out of the total number of threads/posts positively predicted for a given label.

$$\text{Precision} = \frac{TruePositives}{TruePositives + FalsePositives} \tag{2}$$

Recall is defined as the percentage of threads/posts that are correctly and positively categorised by the model out of the total number of threads/posts it should have predicted for that given label.

$$\text{Recall} = \frac{TruePositives}{TruePositives + FalseNegatives} \tag{3}$$

The datasets analysed for our work are highly unbalanced because the majority of posts within them are not investment scam related. Therefore, to complement these performance measures, we also used the F-measure which is a weighted average of precision and recall.

$$\text{F} = 2 \cdot \left(\frac{Precision \cdot Recall}{Precision + Recall} \right) \tag{4}$$

3.4.3. Scam thread type modelling. We test and compare several multi-class text classification methods using supervised, semi-supervised, and unsupervised learning. The feature set used in all models includes the board name, thread title and content of each post. We perform hyperparameter tuning and select the best performing model (according to subsubsection 3.4.2) to classify all posts collected from Bitcointalk and Reddit.

Figure 1 shows the distribution of scam thread type labels in the annotation sample. We can observe that the majority of threads are not related to cryptocurrency investment scams.

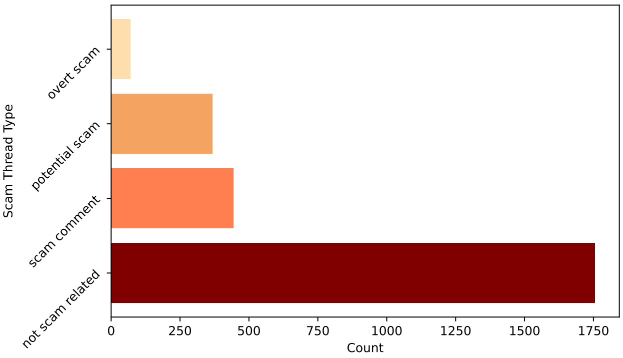


Figure 1: Scam thread type count - BTCtalk annotation sample

We start our analysis by using the XGBoost [42] model as our baseline. We then compare this against three deep learning models that have been used for text classification in other domains [30], [35], [51], namely LSTM [23], CNN-LSTM (a hybrid of CNN [29] and LSTM) and Zero-Shot learning [28].

Zero-shot text classification [10], [28] aims to assign a label to a text document, independently of the text domain and other features implied in the label. Some researchers [58] have used zero-shot models as an unsupervised learning benchmark for text classification. We choose to test this model's performance against other models that do require labelled data.

We also test several versions of LSTM-GloVe using ten different pre-trained word vectors from Stanford-NLP. GloVe [41] is an unsupervised learning algorithm used to obtain words' vector representations. This method uses a global matrix to record the frequency of a word's co-occurrence with one another word in a given corpus. It then uses this matrix for training and the output produces word representations that show linear substructures of the word vector space. For our work, we use LSTM and test word vectors pre-trained on the following datasets[2]:

1) Wikipedia 2014 + Gigaword 5: This dataset contains 6B tokens and 400K vocabularies. It is uncased and has four variations of dimensional vectors: 50d, 100d, 200d, 300d vectors.
2) Common Crawl 42B. This dataset has 42B tokens and 1.9M vocabularies. It is uncased and has 300 dimensional vectors.
3) Common Crawl 840B. This dataset has 840B tokens and 2.2M vocabularies. It is cased and has 300 dimensional vectors.
4) Twitter. This dataset has 2B tweets, 27B tokens and 1.2M vocabularies. It is uncased and has four versions of dimensional vectors: 25d, 50d, 100d, 200d vectors.

We select the method with the highest accuracy and F-measure and implement active learning to test whether these performance measures can be improved. Active learning [47] is a semi-supervised learning method that is iterative and can help improve a classifier's performance. This process makes use of an oracle (or human with domain knowledge) to annotate and add additional labelled data into the training set. Each iteration ranks and obtains the most informative unlabelled inputs so they can be labelled by the oracle.

We select the best performing model to categorise all 531,356 threads extracted from Bitcointalk and all 2,108 submissions collected from Reddit. The classification by thread type helps us identify whether they are investment scam advertisements (overt scams and potential scams), scam comments or not investment scam related.

3.4.4. Scam actor type modelling. Figure 2 shows the distribution of scam actor types in the annotation sample. We can see that 47.45% of actors in the sample are scam owners and 36.55% are scam reporters. To avoid having a low recall score, and poorer subsequent model predictions, we decide to combine scam shills and scam participants since these categories share some similarities and have the lowest count.

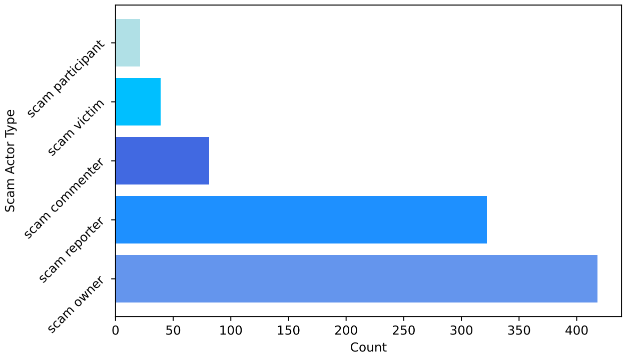


Figure 2: Scam actor type count - BTCtalk annotation sample

We again use the XGBoost [42] model as our baseline. We then implement active learning into the model's pipeline so we can evaluate whether the performance of the classifier can be improved.

We select the best performing model to categorise all posts from Bitcointalk and Reddit identified as potential scams and overt scams as specified in subsubsection 3.4.3. This classification helps us identify which actor type is behind the cryptocurrency investment scam related posts.

3.4.5. Scam lure type modelling. We use our typology inspired by Stajano and Wilson [48] and multi-label classification, as many lures can be used within one cryptocurrency investment scam advert. Figure 3 shows the distribution of scam lure types in the annotation sample classified as per subsection 3.2. The financial principle is found in the majority of annotated posts (95%). This is followed by the distraction and the authority principles which are present in 26% and 20% of the annotation sample respectively. The dishonesty principle is found in 10% and the herd and kindness principles both appear in 7% of all posts. The kindness principle was only identified in 3% of instances.

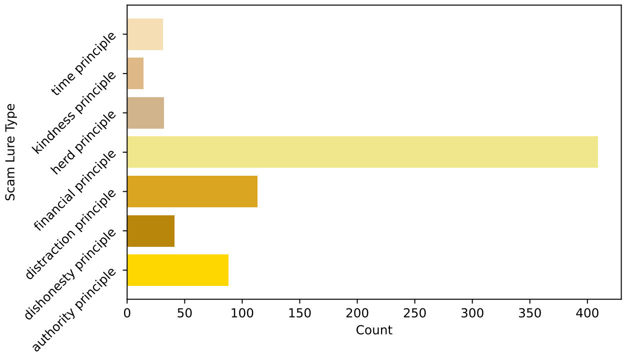


Figure 3: Scam lure type count - BTCtalk annotation sample

We use multiple binary classifiers to identify whether each of the lures defined in subsection 3.2 are present in potential scams and overt scams. We compare the logistic regression model against XGBoost and identify which of these methods have the best performance and produces the best predictions.

4. Results

4.1. Prediction of investment scam advertisements

Table 1 shows the performance results for the initial four models mentioned in 3.4.3. We find our baseline (XGBoost) outperforms the rest of the models with the highest accuracy and F-measure. Out of the four models, it also has fewer false positives and false negatives. The second best performing model is LSTM. We note that our zero-shot learning model has the worst performance of all models.

Table 2 shows the performance of LSTM with all variations for GloVe word embeddings. We can see that the model with highest accuracy and F-measure uses the pre-trained version of GloVe with 27B tokens from Twitter and 100 dimensional vectors. This version outperforms

[2] https://github.com/stanfordnlp/GloVe

TABLE 1: Performance of models for scam thread type

Model	Precision	Recall	F-Measure	Accuracy
XGBoost	0.8167	0.8243	0.8157	0.8243
LSTM	0.7274	0.7405	0.7212	0.7405
CNN-LSTM	0.6859	0.7060	0.6922	0.7061
Zero-shot	0.5592	0.5061	0.5313	0.5061

TABLE 2: Performance of LSTM models with GloVe embeddings for scam thread type

Model	Precision	Recall	F-Measure	Accuracy
Wikipedia 6B 50d	0.7502	0.7646	0.7397	0.7646
Wikipedia 6B 100d	0.7271	0.7324	0.7273	0.7324
Wikipedia 6B 200d	0.7570	0.7646	0.7554	0.7646
Wikipedia 6B 300d	0.7817	0.7910	0.7844	0.7910
Common Crawl 42B 300d	0.7454	0.7451	0.7435	0.7451
Common Crawl 840B 300d	0.7435	0.7497	0.7345	0.7497
Twitter 27B 25d	0.7277	0.7554	0.7332	0.7554
Twitter 27B 50d	0.7539	0.7669	0.7475	0.7669
Twitter 27B 100d	0.7921	0.8036	0.7957	0.8036
Twitter 27B 200d	0.7466	0.7577	0.7448	0.7577

pre-trained versions of GloVe with much larger versions of tokens such as Common Crawl 42B and Common Crawl 840B. We believe that this could be explained by the similarity between Twitter posts and those found on Bitcointalk.

As mentioned in subsubsection 3.4.3, we select the best performing model and implement active learning. Figure 4 and Table 3 show the changes in accuracy during the active learning implementation and the performance results respectively. We can observe that all performance measures show an increase between 0.14% and 0.24%.

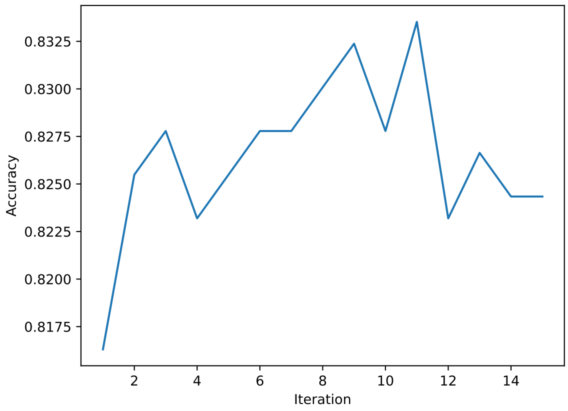


Figure 4: Accuracy changes for XGBoost model with Active Learning - scam thread type

TABLE 3: Performance of XGBoost models for scam thread type

Model	Precision	Recall	F-Measure	Accuracy
XGBoost	0.8167	0.8243	0.8157	0.8243
XGBoost AL	0.8187	0.8255	0.8177	0.8255

Figure 5 displays the confusion matrices for the four best performing models. Figure 5a and Figure 5b show that XGBoost and XGBoost with active learning produce the fewest false positives and false negatives.

We use the XGBoost with active learning model to categorise all 531,356 threads extracted from Bitcointalk and all 2,108 submissions collected from Reddit. The classification by thread type helps us identify whether they are investment scam advertisements (overt scams and potential scams), scam comments or not investment scam related. Table 4 and Table 5 show the prediction results for Bitcointalk and Reddit respectively.

TABLE 4: Bitcointalk thread type predictions with XGBoost model

Thread type	Number of threads	Percentage
Not scam related	498,911	93.89%
Potential scam	22,296	4.20%
Scam comment	9,586	1.80%
Overt scam	563	0.11%
Total count	531,356	100.00%

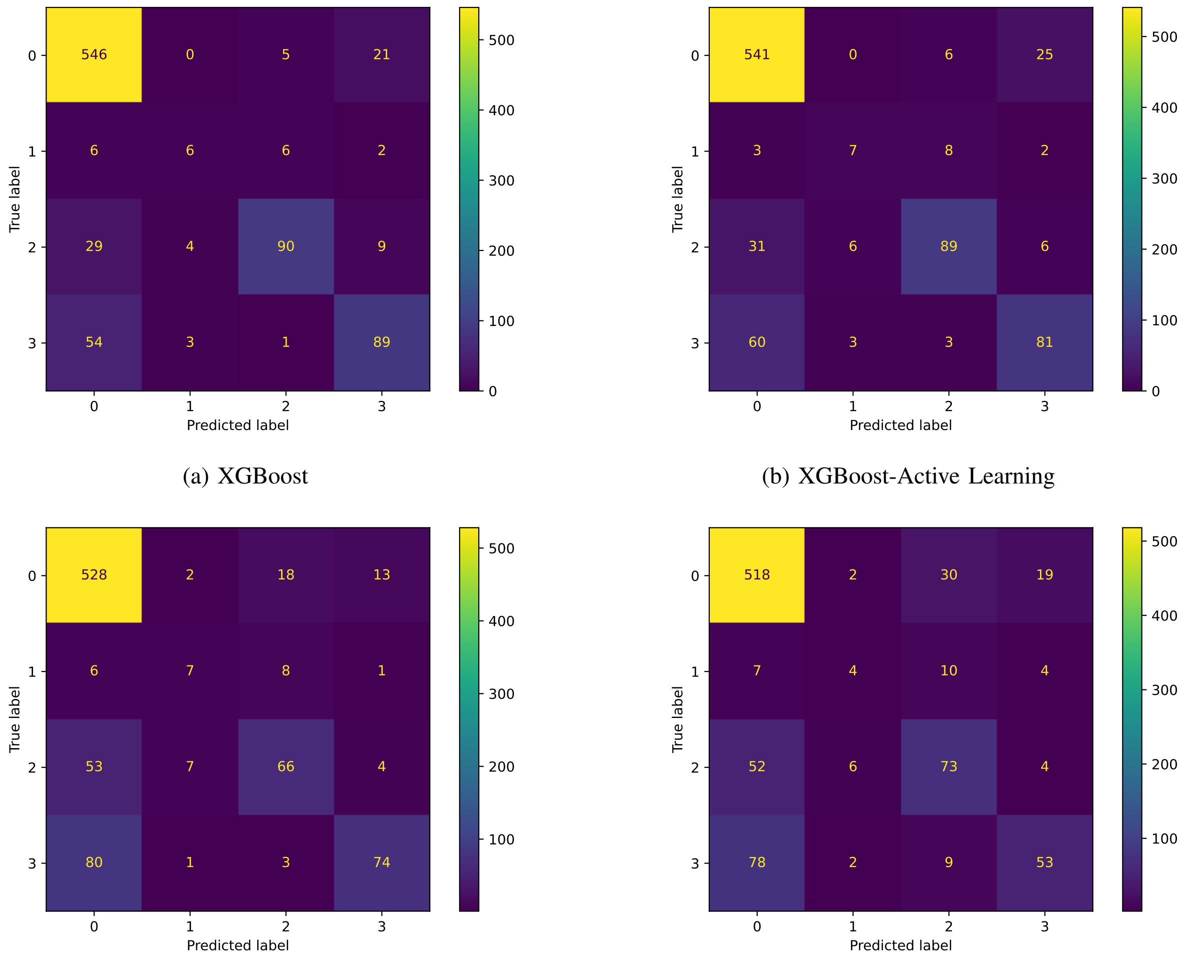
As expected, we can observe the majority of predictions (93.89% of Bitcointalk threads and 93.74% of Reddit posts) are not cryptocurrency investment scam related. The categorisation results of potential scams are also consistent between both platforms since we uncover that around five percent of all posts collected on both sites belong to this category (4.20% of Bitcointalk threads and 5.88% of Reddit posts). We note there are no predictions of overt scams found within Reddit. This is compatible with our expectations since this platform is heavily moderated by its administrators and it has dedicated subreddits for topics that are Ponzi and HYIP related (for example, r/PonziSchemes, r/DefiPonzi, r/HYIP, r/HYIPCommunity, r/HyipMonitors, etc.).

TABLE 5: Reddit thread type predictions with XGBoost model

Thread type	Number of threads	Percentage
Not scam related	1,976	93.74%
Potential scam	124	5.88%
Scam comment	8	0.38%
Total count	2,108	100.00%

4.2. Prediction of actors involved in investment scam advertisements

As mentioned in subsubsection 3.4.4, we implement active learning with the XGBoost model to check for changes in performance measures. Figure 6 and Table 6 show the changes in accuracy during the active learning implementation and the performance results respectively. Surprisingly, we observe that all performance measures show a decrease between 1.6% and 2.2% using active learning.



(a) XGBoost

(b) XGBoost-Active Learning

(c) LSTM-GloVe Twitter 27B 100d

(d) LSTM-GloVe Wikipedia 6B 300d

Figure 5: Confusion matrices for the best four scam thread type models
0 - Not investment scam related, 1 - Overt scam, 2 - Potential scam, 3 - Scam comment

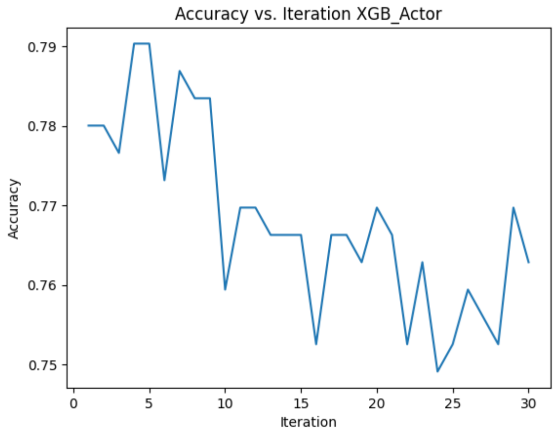


Figure 6: Accuracy changes for XGBoost model with Active Learning - scam actor type

TABLE 6: Performance of statistical models for scam actor type

Model	Precision	Recall	F-Measure	Accuracy
XGBoost	0.6944	0.7801	0.7139	0.7801
XGBoost AL	0.6832	0.7629	0.7022	0.7629

As active learning in this case has not helped increase the performance of the XGBoost model, we decide to use the XGBoost model without active learning to classify all potential and overt scams. Figure 7 shows the confusion matrix for the XGBoost model.

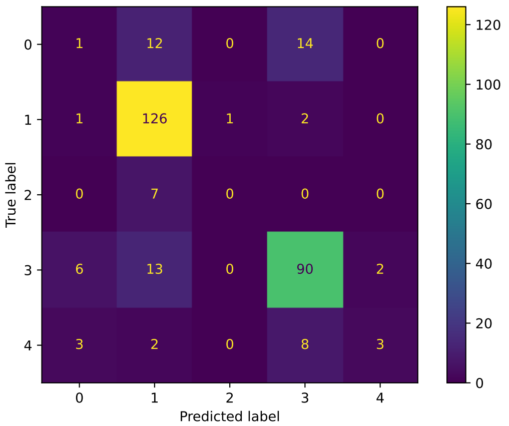


Figure 7: Confusion matrix - XGBoost model for actor type
0 - scam commenter, 1 - scam owner, 2 - scam victim, 3 - scam reporter, 4 - scam participant

We categorise all 22,859 scam related threads (potential scams and overt scams) from Bitcointalk and 124 potential scams from Reddit. Table 7 and Table 8 show the prediction results for Bitcointalk and Reddit respectively. We can see that the majority of actors (96.82% of Bitcointalk and 97.58% of Reddit) are scam owners.

TABLE 7: Scam actor type predictions with XGBoost model on Bitcointalk

Actor type	Number of threads	Percentage
Scam owner	22,132	96.82%
Scam reporter	520	2.27%
Scam commenter	172	0.75%
Scam victim	3	0.01%
Scam participant	32	0.14%
Total count	22,859	100.00%

TABLE 8: Scam actor type predictions with XGBoost model on Reddit

Actor type	Number of threads	Percentage
Scam owner	121	97.58%
Scam commenter	3	2.42%
Total count	124	100.00%

4.3. Prediction of lure types used in investment scam advertisements

Table 9 and Table 10 show the performance results for predicting lure type using logistic regression and XGBoost, as outlined in 3.4.5. We observe that the logistic regression models and XGBoost models have very similar performance for the financial, herd, kindness and time principles. On the other hand, the XGBoost models for authority, dishonesty and distraction principles outperform the logistic regression models.

We use all logistic regression and XGBoost models to identify lure principles used in all 22,859 scam related threads (potential scams and overt scams) from Bitcointalk and 124 potential scams from Reddit. Table 11 shows the prediction results for scam lure types used in Bitcointalk and Reddit using XGBoost. While the performance measures for the XGBoost and logistic regression models are within similar ranges (between 0.68 and 0.9 for F-measure and 0.75 and 0.97 for accuracy), the logistic regression model provides poor predictions. The XGBoost model predictions are consistent with the training data. Therefore, we focus our analysis on this models' predictions. We observe that the financial principle is present in the majority of scam related posts in both platforms. We notice some differences between the second and third most popular lures. In Bitcointalk, the second most popular lure type is the distraction principle, (appearing in 36.12% of scam adverts), followed by the authority principle (present in 14.53% of scam adverts). On the other hand, scam lure predictions on Reddit show that the authority principle appears more frequently (20.16% of predictions) than the distraction principle (6.45% of predictions).

TABLE 9: Performance of logistic regression models for each scam lure type

Model	Precision	Recall	F-Measure	Accuracy
Authority principle	0.6110	0.7817	0.6859	0.7817
Dishonesty principle	0.8381	0.9155	0.8751	0.9155
Distraction principle	0.8704	0.8732	0.8681	0.8732
Financial principle	0.8381	0.9155	0.8751	0.9155
Herd principle	0.7999	0.8944	0.8445	0.8944
Kindness principle	0.9308	0.9648	0.9475	0.9648
Time principle	0.9720	0.9859	0.9789	0.9859

TABLE 10: Performance of XGBoost models for each scam lure type

Model	Precision	Recall	F-Measure	Accuracy
Authority principle	0.7085	0.7535	0.7220	0.7535
Dishonesty principle	0.8914	0.9155	0.8956	0.9155
Distraction principle	0.9071	0.9085	0.9063	0.9085
Financial principle	0.8381	0.9155	0.8751	0.9155
Herd principle	0.7999	0.8944	0.8445	0.8944
Kindness principle	0.9308	0.9648	0.9475	0.9648
Time principle	0.9718	0.9718	0.9718	0.9718

TABLE 11: Scam lure type predictions with XGBoost models on Bitcointalk and Reddit

	Bitcointalk		Reddit	
Scam Lure type	Count	Percentage	Count	Percentage
Authority principle	3,322	14.53%	25	20.16%
Dishonesty principle	140	0.61%	0	0.00%
Distraction principle	8,257	36.12%	8	6.45%
Financial principle	22,839	99.91%	124	100.00%
Herd principle	21	0.09%	0	0.00%
Kindness principle	29	0.13%	0	0.00%
Time principle	170	0.74%	1	0.81%
Total count	22,859	100.00%	124	100.00%

4.4. Scam advertisements targeted at pensioners

Our final objective is to understand whether pensioners are targeted more heavily with investment scam advertisements than other investor communities. Table 12 shows the classification results for each of the four subreddits selected. We can see the subreddit r/investment has the highest proportion of potential scams (25.91%) followed by r/InvestmentClub (13.89%) which also has the highest ratio of scam comments (2.78%). Subreddits r/investing and r/retirement have a very similar distribution of potential scams and scam comments. These results indicate that members of r/retirement are not being targeted with a higher number of cryptocurrency investment scam adverts.

5. Conclusion

The objectives of this research are to analyse and compare different machine learning models to identify cryptocurrency investment scam adverts with high accuracy in the online forum Bitcointalk and the social media platform Reddit. An additional aim is to identify whether people that have reached a retirement age, and therefore coming into possession of a pension pot, are being targeted more frequently than other investors by criminals through cryptocurrency investment scam advertisements.

We built three text classifiers and compared several models using supervised, semi-supervised and unsupervised learning. Our results show that the baseline for the scam thread type classifier, the XGBoost model, outperformed three other models that have been used for multiclass text classification in other domains, namely, LSTM, CNN-LSTM and Zero-Shot learning. While there are benefits to using Zero-Shot learning, in that the unsupervised approach lowers the overhead for manual annotations, in our test its performance was quite poor.

We also tested the LSTM model with ten different versions of pre-trained GloVe word embeddings. We identified that the dataset pre-trained with Twitter 27B tokens and 100 dimensional vectors outperforms datasets with higher number of tokens and dimensional vectors. We believe that this is related to the text similarities found between our data sources and Twitter posts.

The XGBoost model implemented with active learning had the best performance measures overall and was used to categorise all posts extracted from both data sources. Our categorisation by scam thread type showed that 4.2% and 5.88% of our Bitcointalk and Reddit posts, respectively, are potential scams. We also found that 0.11% of all Bitcointalk posts are overt scams and 1.8% are scam comments. Our classifier did not find any overt scams on the posts extracted from Reddit. We believe that this supports the accuracy of our scam thread classifier since this platform has specific subreddits where submissions about Ponzi schemes and HYIPs are posted.

We identify predictions of cryptocurrency investment scam adverts (potential and overt scams) in both platforms (22,859 in Bitcointalk and 124 in Reddit) and use the XGBoost model as the scam actor classifier. We identify that the majority of actors on both platforms are scam owners. We then use logistic regression and XGBoost models to build binary classifiers for scam lure types. Our results show that even though these two types of models have similar performance measures, the XGBoost models provide better predictions of scam lure types used in cryptocurrency scam advertisements.

We find the financial principle to be present in the majority of potential and overt scams. The distraction principle, where the post is overloaded with irrelevant detail, is found in over one-third of scam lures found on Bitcointalk, but only 6% on Reddit lures. On Reddit, the authority principle, where fraudsters leverage technical and authoritative references, is used in about 20% of scam lures. Investors in general should be aware of these lures when they use online forums and social media.

Our thread type classification results of four subreddits related to investment and retirement shows that only 2.94% of potential scams are found in the subreddit r/retirement. This percentage is smaller than the proportion of potential scams in the subreddits r/investment and r/InvestmentClub and similar to r/investing. Therefore, we infer that pensioners are not targeted more heavily through adverts of cryptocurrency scams through Reddit posts. An alternative explanation is that the subreddit is well-moderated, and scam invitations are quickly removed. It could also be possible that those scam adverts are only allowed to be posted in dedicated subreddits related to Ponzi schemes and HYIPs (r/PonziSchemes, r/DefiPonzi, r/HYIP, r/HYIPCommunity, r/HyipMonitors, etc.) which can be accessed by any type of Reddit user.

Finally, we should note that scammers can take advantage of the economic challenges and fears that pensioners and investors in general face when managing their lifetime savings [50]. These users must be aware that social media platforms and online forums do contain cryptocurrency investment scam advertisements. They also should be familiar with the different types of lures used by cybercriminals so they do not fall prey of fraudulent schemes.

6. Acknowledgments

This work is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 949127, for GAS and AH) and the Engineering and Physical Sciences Research Council (EPSRC) (grant agreement No EP/W032473/1, for AH).

As part of the open-report model followed by the Workshop on Attackers & CyberCrime Operations (WACCO), all the reviews for this paper are publicly available at https://github.com/wacco-workshop/WACCO/tree/main/WACCO-2023.

References

[1] British Society of Criminology. Statement of Ethics. https://www.britsoccrim.org/ethics/, 2015.

[2] Sharad Agarwal, Gilberto Atondo-Siu, Marilyne Ordekian, Alice Hutchings, Enrico Mariconti, and Marie Vasek. Short paper: Defi deception–uncovering the prevalence of rugpulls in cryptocurrency projects. Financial Cryptography and Data Security, 2023.

[3] Gilberto Atondo Siu, Alice Hutchings, Marie Vasek, and Tyler Moore. "invest in crypto!": An analysis of investment scam advertisements found in bitcointalk. APWG, 2022.

[4] Emad Badawi and Guy-Vincent Jourdan. Cryptocurrencies emerging threats and defensive mechanisms: A systematic literature review. *IEEE Access*, 8:200021–200037, 2020.

[5] Emad Badawi, Guy-Vincent Jourdan, Gregor Bochmann, and Iosif-Viorel Onut. An automatic detection and analysis of the Bitcoin generator scam. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 407–416. IEEE, 2020.

TABLE 12: Reddit thread type predictions with XGBoost model

	Subreddit							
	r/investment		**r/InvestmentClub**		**r/investing**		**r/retirement**	
Thread type	Count	Percentage	Count	Percentage	Count	Percentage	Count	Percentage
Not scam related	159	72.27%	90	83.33%	901	96.99%	826	97.06%
Potential scam	57	25.91%	15	13.89%	27	2.91%	25	2.94%
Scam comment	4	1.82%	3	2.78%	1	0.11%	0	0.00%
Total count	220	100.00%	108	100.00%	929	100.00%	851	100.00%

[6] Massimo Bartoletti, Salvatore Carta, Tiziana Cimoli, and Roberto Saia. Dissecting ponzi schemes on ethereum: identification, analysis, and impact. *Future Generation Computer Systems*, 102:259–277, 2020.

[7] Daniele Bianchi. Cryptocurrencies as an asset class? an empirical assessment. *The Journal of Alternative Investments*, 23(2):162–179, 2020.

[8] Yazan Boshmaf, Charitha Elvitigala, Husam Al Jawaheri, Primal Wijesekera, and Mashael Al Sabah. Investigating MMM P†onzi scheme on bitcoin. In *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security*, pages 519–530, 2020.

[9] Alexandra Burton, Claudia Cooper, Ayesha Dar, Lucy Mathews, and Kartikeya Tripathi. Exploring how, why and in what contexts older adults are at risk of financial cybercrime victimisation: A realist review. *Experimental Gerontology*, 159:111678, 2022.

[10] Ming-Wei Chang, Lev-Arie Ratinov, Dan Roth, and Vivek Srikumar. Importance of semantic representation: Dataless classification. In *Aaai*, volume 2, pages 830–835, 2008.

[11] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, Jun 2002.

[12] W. Chen, Z. Zheng, E. C. . Ngai, P. Zheng, and Y. Zhou. Exploiting blockchain data to detect smart ponzi schemes on ethereum. *IEEE Access*, 7:37575–37586, 2019.

[13] Weili Chen, Zibin Zheng, Jiahui Cui, Edith Ngai, Peilin Zheng, and Yuren Zhou. Detecting ponzi schemes on ethereum: Towards healthier blockchain technology. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, page 1409–1418, 2018.

[14] Shaen Corbet, Andrew Meegan, Charles Larkin, Brian Lucey, and Larisa Yarovaya. Exploring the dynamic relationships between cryptocurrencies and other financial assets. *Economics Letters*, 165:28–34, 2018.

[15] Marco L. Della Vedova, Eugenio Tacchini, Stefano Moret, Gabriele Ballarin, Massimo DiPierro, and Luca de Alfaro. Automatic online fake news detection combining content and social signals. In *2018 22nd Conference of Open Innovations Association (FRUCT)*, pages 272–279, 2018.

[16] Jake Drew and Tyler Moore. Automatic identification of replicated criminal websites using combined clustering. In *2014 IEEE Security and Privacy Workshops*, pages 116–123, 2014.

[17] Anne Haubo Dyhrberg. Bitcoin, gold and the dollar – a garch volatility analysis. *Finance Research Letters*, 16:85–92, 2016.

[18] Stefan Ehlers and Kolja Gauer. Beyond bitcoin: A statistical comparison of leading cryptocurrencies and fiat currencies and their impact on portfolio diversification. *The Journal of Alternative Investments*, 22(1):114–125, 2019.

[19] European Commission. Proposal for a regulation of the european parliament and of the council on markets in crypto-assets, and amending directive (eu) 2019/1937. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020PC0593, 2020.

[20] FBI.GOV. Federal Bureau of Investigation - Investment Crime Report. https://www.ic3.gov/Media/PDF/AnnualReport/2022_IC3Report.pdf, 2022.

[21] J.T. Hamrick, Farhang Rouhi, Arghya Mukherjee, Amir Feder, Neil Gandal, Tyler Moore, and Marie Vasek. An examination of the cryptocurrency pump-and-dump ecosystem. *Information Processing & Management*, 58(4):102506, 2021.

[22] Jochen Hartmann, Juliana Huppertz, Christina Schamp, and Mark Heitmann. Comparing automated text classification methods. *International Journal of Research in Marketing*, 36(1):20–38, 2019.

[23] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9:1735–80, 12 1997.

[24] Giacomo Ibba, Giuseppe Antonio Pierro, and Marco Di Francesco. Evaluating machine-learning techniques for detecting smart ponzi schemes. In *2021 IEEE/ACM 4th International Workshop on Emerging Trends in Software Engineering for Blockchain (WETSEB)*, pages 34–40, 2021.

[25] IC3.gov. Elder fraud report 2021. https://www.ic3.gov/Media/PDF/AnnualReport/2021_IC3ElderFraudReport.pdf, 2021.

[26] Abdul Azim Ismail and Marina Yusoff. An efficient hybrid lstm-cnn and cnn-lstm with glove for text multi-class sentiment classification in gender violence. *International Journal of Advanced Computer Science and Applications*, 13(9), 2022.

[27] Nir Kshetri. Scams, frauds, and crimes in the nonfungible token market. *Computer*, 55(4):60–64, 2022.

[28] Hugo Larochelle, Dumitru Erhan, and Yoshua Bengio. Zero-data learning of new tasks. In *AAAI*, volume 1, page 3, 2008.

[29] Yann LeCun et al. Generalization and network design strategies. *Connectionism in Perspective*, 19(143-155):18, 1989.

[30] Hongxia Lu, Louis Ehwerhemuepha, and Cyril Rakovski. A comparative study on deep learning models for text classification of unstructured medical notes with various levels of class imbalance. *BMC Medical Research Methodology*, 22(1):181, 2022.

[31] Simon Mackenzie. Criminology towards the metaverse: Cryptocurrency scams, grey economy and the technosocial. *British Journal Of Criminology*, 2022.

[32] Omri Marian. A conceptual framework for the regulation of cryptocurrencies. *U. Chi. L. Rev. Dialogue*, 82:53, 2015.

[33] Nigel Martin and John Rice. Spearing high net wealth individuals: the case of online fraud and mature age internet users. *International Journal of Information Security and Privacy (IJISP)*, 7(1):1–15, 2013.

[34] Bruno Mazorra, Victor Adan, and Vanesa Daza. Do not rug on me: Leveraging machine learning techniques for automated scam detection. *Mathematics*, 10(6), 2022.

[35] Blake Miller, Fridolin Linder, and Walter R Mebane. Active learning approaches for labeling text: review and assessment of the performance of active learning approaches. *Political Analysis*, 28(4):532–551, 2020.

[36] Tyler Moore, Jie Han, and Richard Clayton. The postmodern Ponzi scheme: Empirical analysis of high-yield investment programs. In *Financial Cryptography and Data Security*, volume 7397 of *Lecture Notes in Computer Science*, pages 41–56. Springer, 2012.

[37] Jens Neisius and Richard Clayton. Orchestrated crime: The high yield investment fraud ecosystem. In *2014 APWG Symposium on Electronic Crime Research (eCrime)*, volume 2014-, pages 48–58. IEEE, 2014.

[38] James Nicholson, Lynne Coventry, and Pamela Briggs. ”if it's important it will be a headline” cybersecurity information seeking in older adults. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2019.

[39] OECD. Why decentralised finance (defi) matters and the policy implications, 2022.

[40] Sergio Pastrana, Daniel R. Thomas, Alice Hutchings, and Richard Clayton. CrimeBB: Enabling cybercrime research on underground forums at scale. In *Proceedings of the 2018 World Wide Web Conference*, WWW '18, page 1845–1854, 2018.

[41] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, 2014.

[42] Reena Shaw. Xgboost: A concise technical overview, www.kdnuggets.com/2017/10/xgboost-concise-technical-overview, 2017.

[43] Niranjan Sapkota, Klaus Grobys, and Josephine Dufitinema. How much are we willing to lose in cyberspace? on the tail risk of scam in the market for initial coin offerings. https://ssrn.com/abstract=3732747, 2020.

[44] Sarah Sarabadani, Gaurav Baruah, Yan Fossat, and Jouhyun Jeon. Longitudinal changes of covid-19 symptoms in social media: Observational study. *J Med Internet Res*, 24(2):e33959, Feb 2022.

[45] SEC's Office of Investor Education and Advocacy. Investor alert - ponzi schemes using virtual currencies. https://www.sec.gov/files/ia_virtualcurrencies.pdf, 2023.

[46] Semrush. Reddit traffic. https://www.semrush.com/website/reddit.com/overview/, 2023.

[47] Burr Settles. Active learning literature survey. *University of Wisconsin-Madison Department of Computer Sciences*, 2009.

[48] Frank Stajano and Paul Wilson. Understanding scam victims: seven principles for systems security. *Communications of the ACM*, 54(3):70–75, 2011.

[49] Statista. Number of identity-verified cryptoasset users from 2016 to November 2022 (in millions). https://www.statista.com/statistics/1202503/global-cryptocurrency-user-base/, 2022.

[50] The Pensions Regulator. Scam-fighting bodies tell pension savers to stay on guard and get guidance. https://www.thepensionsregulator.gov.uk/en/media-hub/press-releases/2022-press-releases/scam-fighting-bodies-tell-pension-savers-to-stay-on-guard-and-get-guidance, 2022.

[51] Santosh Tokala, Vaibhav Gambhir, and Animesh Mukherjee. Deep learning for social media health text classification. In *Proceedings of the 2018 EMNLP Workshop SMM4H: The 3rd Social Media Mining for Health Applications Workshop & Shared Task*, pages 61–64, Brussels, Belgium, October 2018. Association for Computational Linguistics.

[52] K. Toyoda, P. Takis Mathiopoulos, and T. Ohtsuki. A novel methodology for HYIP operators' bitcoin addresses identification. *IEEE Access*, 7:74835–74848, 2019.

[53] Arianna Trozze, Josh Kamps, Eray Arda Akartuna, Florian J Hetzel, Kleinberg Bennett, Toby Davies, and Shane D Johnson. Cryptocurrencies and future financial crime. *Crime Science*, 11(1), 2022.

[54] Marie Vasek and Tyler Moore. There's no free lunch, even using bitcoin: Tracking the popularity and profits of virtual currency scams. volume 8975 of *Lecture Notes in Computer Science*, pages 44–61, Berlin, Heidelberg, 2015. Springer Berlin Heidelberg.

[55] Marie Vasek and Tyler Moore. Analyzing the Bitcoin Ponzi scheme ecosystem. In *Fifth Workshop on Bitcoin and Blockchain Research*, Lecture Notes in Computer Science. Springer, 2018.

[56] Kristin Weber, Andreas E. Schütz, Tobias Fertig, and Nicholas H. Müller. Exploiting the human factor: Social engineering attacks on cryptocurrency users. In Panayiotis Zaphiris and Andri Ioannou, editors, *Learning and Collaboration Technologies. Human and Technology Ecosystems*, pages 650–668, Cham, 2020. Springer International Publishing.

[57] Pengcheng Xia, Haoyu Wang, Xiapu Luo, Lei Wu, Yajin Zhou, Guangdong Bai, Guoai Xu, Gang Huang, and Xuanzhe Liu. Don't fish in troubled waters! characterizing coronavirus-themed cryptocurrency scams. In *2020 APWG Symposium on Electronic Crime Research (eCrime)*, pages 1–14, 2020.

[58] Wenpeng Yin, Jamaal Hay, and Dan Roth. Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3914–3923, Hong Kong, China, November 2019. Association for Computational Linguistics.

[59] Fatima Zahrah, Jason RC Nurse, and Michael Goldsmith. A comparison of online hate on reddit and 4chan: a case study of the 2020 us election. In *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing*, pages 1797–1800, 2022.

[60] Linda Zhou, Andrew Caines, Ildiko Pete, and Alice Hutchings. Automated hate speech detection and span extraction in underground hacking and extremist forums. *Natural Language Engineering*, page 1–28, 2022.