Twitter Sentiment Analysis and Expert Ratings of Initial Coin Offering Fundraising: Evidence from Australia and Singapore Markets

Anchaya Chursook, Ahmad Yahya Dawod, Somsak Chanaim, Nathee Naktnasukanjn, Nopasit Chakpitak

International College of Digital Innovation, Chiang Mai University, 239 Nimmanahaeminda Road, Suthep, Muang, Chiang Mai 50200, Thailand

***Abstract* – Sentiment analysis of Twitter data is quite valuable for determining the market opinion. Twitter sentiment analysis is more challenging than generic sentiment analysis owing to slang and misspellings. The techniques utilized for evaluating the sentiment of tweets that have the greatest importance for the success of an Initial Coin Offering (ICO) are machine learning approaches. In this study, we examined market sentiment and used Expert Ratings to predict the success of ICOs in the Australian and Singapore markets. Based on 68,281 tweets from 57 ICOs across four industries: business services, cryptocurrency, entertainment, and platform. Several classification methods were investigated, including Support Vector Machines (SVMs), Logistic Regression (LR), Random Forest (RF), and Naïve Bayes (NB). The outcomes indicated that sentiment analysis of tweets and expert ratings may be used to forecast the success of an initial coin offering. The results indicate that the suggested model is capable of accurately assessing the tweets of the ICO Successful with a maximum accuracy of about 94.7 % when implementing the Support Vector Machines (SVMs) classifier.**

DOI: 10.18421/TEM111-06
https://doi.org/10.18421/TEM111-06

Corresponding author: Ahmad Yahya Dawod,
International College of Digital Innovation, Chiang Mai University. Chiang Mai 50200, Thailand
Email: ahmadyahyadawod.a@cmu.ac.th

Received: 19 October 2021.
Revised: 16 January 2022.
Accepted: 21 January 2022.
Published: 28 February 2022.

***Keywords* – Sentiment Analysis, Initial Coin Offerings, Natural Language Processing, Opinion Mining, Tweets.**

1. Introduction

Crowdfunding is the new fundraising method related to blockchain technology and cryptocurrency such as Initial Coin Offering (ICO) and Initial Exchange Offering (IEO). The ICOs, also known as token sales or crowd sales, are a smart contract-based fundraising technique for organizations, companies, and entrepreneurs [1]. In the business world, the ICOs are popular because they provide a quick, low-cost, and easily accessible option to raise capital. The ICOs' issuers utilize blockchain technology to generate new coins, which they then sell on their website. Social media platforms are being utilized to generate and spread awareness among international investors [2]. The IEOs are managed by cryptocurrency exchanges platform on behalf of the issuing business in order to raise funds through the selling of new tokens issuers [3]. In 2019, there have already been 32 IEOs launched on the website ICObench.com. The completed IEOs have raised over $159 million. The cryptocurrency exchanges act as intermediaries between projects and investors by providing crowdsale services [4], which include the sale of tokens to individual contributors, among other things. Through this funding, such initiatives can save money while maintaining security and dependability, as well as allowing investors to put their money into high-quality projects.

The ICO fundraising is a popular method of funding for startups. The ICOs have raised a significant amount of money, especially in 2017 and 2018. In 2018, a total of 3,782 ICOs were launched, raising almost $11.4 billion [5], [6]. In 2021, 113 ICOs have already been issued on the platform of coincodex. Digital tokens are being created with the use of blockchain technology, and these digital tokens may be produced rapidly, reliably, and instantaneously using smart contracts [7]. The success of the ICO fundraising campaign due to the concept, and it takes time to establish and gain market confidence. A whitepaper for an initial coin offering is a pre-announcement of the project. The whitepaper provides public information about the firm and describes the business strategy, blockchain platform, token sale supply, token sale structure, token distribution, use of money, team, and roadmap, among other things [8]. A lot of popular social media platforms were utilized by the ICOs to promote their offerings, including Telegram, Twitter, Facebook, Reddit, Instagram, Bitcointalk, Medium, and YouTube [9]. Airdrop is one of the marketing strategies used to increase public awareness of new ICOs or IEOs. The Airdrop is a marketing tactic used by businesses to reward participants in marketing activities with free digital tokens or cryptocurrency. For the activities, participants are encouraged to do such tasks as joining a telegram group, following on Twitter, Facebook, Instagram, and then retweeting and tagging friends [10]. In addition, participants are required to subscribe to the YouTube channel. After completing the assignment successfully, you may submit a BEP-20 wallet to receive the free token. This airdrop enables you to launch a completely virtual ICO and utilize it to increase the company's awareness. By boosting community engagement and, as a result, brand endorsement from customers, this method lets more individuals invest in the ICOs [11]. Twitter Sentiment analysis and Expert Ratings of ICOs fundraising in terms of Machine learning approaches are used for evaluating the sentiment of the tweets, which is critical for the success of an Initial Coin Offering (ICO). 2. Related work, 3. Methods and Materials, 4. Result of our model and discussion, 5. Conclusion.

2. Related Work

Sentiment analysis has provided a crucial result to assess how the ICOs are successful. Twitter Comment also provides a practical result to strengthen the voice of the customer to be more competitive in the world market [12]. Many social media channels are being used as a marketing tool to advertise products and connect with potential consumers. Twitter is a prominent social networking website worldwide for tweeting brief messages up to 280 characters and can tweet audio messages for up to 140 seconds with the message provider to connect instantly by tweeting, retweeting, and often by a hashtag.

It can discover what is happening and the most talked-about topics associated with Twitter's currently popular hashtags [13]. Nowadays, digital marketing trends use influencers to tweet about goods and services, since influencers may represent and shape social media attitudes about brand trustworthiness, which results in buying choices. When compared to other channels, Twitter is the most efficient way to drive discussions, with a 4:1 feedback ratio for each conversation started on the platform. A strong Twitter presence may increase brand recognition while also encouraging brand discussion. In addition to assisting marketing funnels, conversations may be used to attract new audiences and broaden the discourse, allowing it to be extremely successful. According to a Twitter study, a 10% increase in product or brand conversations can result in a 3% boost in sales. In this paper, we present a model to predicting the ICO fundraising success with tweets and expert ratings from businesses based in Australia and Singapore. The method is utilizing sentiment analysis of the tweets. The successful ICOs Comments on Twitter related to the ICOs are also collected for data analysis. Sentiment analysis will be applied to the collected data to understand behavior and the social sentiment of investors to the ICOs resulting in marketing strategy establishment [14].

3. Method and Materials

3.1. Proposed Method

The proposed system introduces a novel technique for Sentiment analysis of Twitter data as shown in Figure 1. The ICO Fundraising is our contribution to this research, and we attempt to involve ICO data, ICO Rating, and ICO Tweet.

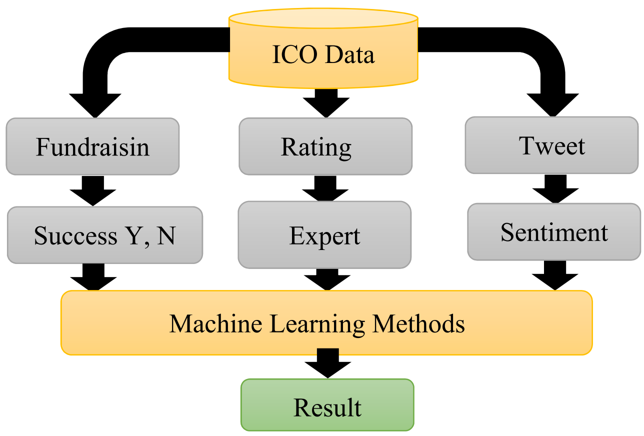

Figure 1. The framework of the Research method

A. ICO Data

We collect the following information on the ICO whitepaper including the Hard Cap, Soft Cap, and collect data from the ICO website to collect data about raised funds, and the ICO expert ratings. In this research, we have studied the ICO market in Australia and Singapore. Asia has been a significant source of ICO funding. Singapore, Hong Kong, and

China are the top three Asian economies in terms of the amount of the ICOs. Although the trend of the ICOs has decreased in 2019, the top three worldwide ICO marketplaces are the United States, Singapore, and England, in that order. Singapore is the world's most competitive economy, ranking fourth, while Australia is ranked sixth in the 2021 Global fintech ranking. The number of firms that have successfully raised capital and are producing income may be used to determine the viability of a FinTech ecosystem. Singapore presently has over 1,000 fintech businesses. Singapore's government is responsible for regulating rules and supporting the development of creative ecosystems for entrepreneurs, such as investment encouragement and legislation. The Monetary Authority of Singapore (MAS) is Singapore's integrated regulator and supervisor of financial institutions, assisting the city-state in its role as Asia's Fintech Center. Australia has risen two spots in the global fintech rankings [15]. The country is now ranked sixth in the world and second in the Asia Pacific. Australia is one of the nations where FinTech has grown considerably and where there is a lot of support for Fintech growth. According to Fintech Australia, the industry has grown from a $250 million industry in 2015 to a $4 billion industry by 2021. Australia is an appealing location for global Fintech investment due to its mature, diverse, and globally connected ecosystem.

The first database we use is the ICO data from the whitepaper and the website www.ico.tokens-economy.com. ICO Whitepapers are the most detailed information sources on the firm offered to ICO investors [16]. This is the most recent overall information about the ICOs, including the business strategy, blockchain platform, token sale supply, token sale structure, token distribution, usage of funds, Hard Cap, Soft Cap, team, and roadmap. The data from www.ico.tokens-economy.com will be used to organize data by year, country, sector, or ICO category. In this step, it is important to analyze the Hard Cap and Soft Cap data. The second database is www.icobench.com. ICObench is a rating tool for the ICOs that is backed by investors and financial professionals. This website is widely recognized as a resource for obtaining the ICOs and the IEO data. It presently provides an overview of ICO information, funds raised, and ICO ratings from blockchain entrepreneurs and fintech firms. In this phase, we compare the amount of fundraising with the Hard Cap and Soft Cap. Lastly, we use tweets on Twitter for sentiment analysis for the prediction of market success. The number of tweets is a sign of the fundraising volume. The relationships between startups' raised funds and Twitter sentiment show increasing emotionality in tweets towards the successful ICO [17].

B. Fundraising

Fundraisers require social marketing, while social marketers require fundraising. Both can now benefit from the DSC's brand-new edition of Marketing Strategy for Effective Fundraising, which includes insights, research, and best practices [18]. Fundraisers at all levels of an organization must understand and apply good Social Marketing concepts in order to develop and implement plans that are both long-term and short-term in nature. Marketers who are good at what they do always try to understand their markets and their audiences. Similarly, fundraisers must understand why people give and what the triggers are behind any charitable behavior if they are to develop and implement strategies to generate long-term income. In this process, we collect the ICOs financial data on www.icobench.com. For each ICO financial data include token information such as token name, platform, type, price in the ICO, average price, a token of sale, and fund raised. We collect the raised fund volume in dollars, the coin's initial price, and the name of the accompanying Twitter account. All ICOs with a raised volume of zero and those that did not create at least one tweet before their ICO date are removed. Outliers are accounted for by excluding the ICOs with a rising volume in the top percentile.

C. Success Y, N

The soft cap is the smallest amount of money that may be raised through a crowd sale campaign to establish a business [19]. The soft cap has such a detrimental impact on the chances of a successful campaign that issuers with a minimum fundraising barrier may fall short of their goal. An ICO will be considered a success if it raises more than its soft cap. In other words, failure is defined as an ICO that fails to obtain enough funds to surpass its soft cap. This part of our analysis examines the successful ICO by comparing the amount of raised funds and soft cap as illustrated in (1). If the ICO can raise funds (R) more than its soft cap (S), the successful ICO Fundraising (X) is equal or greater than 0 and the failure of the ICO Fundraising is less than 0. In our calculations, we substitute "Y" to represent the Successful ICO and "N" to represent the failure ICO.

$$X = \begin{cases} success, R - S \geq 0 \\ fail, \text{else} \end{cases} \tag{1}$$

Where; X is the successful ICO Fundraising, R is the amount of raised funds, S is an ICO soft cap.

D. Expert rating

The next step is to collect the ICO Expert Rating dataset, which is then preprocessed to ensure its reliability and relevance to successful ICOs. In order to test Hypothesis 2, we include the variable ICO Expert Ratings in our model in order to determine the link between the ICO success and the ICO expert ratings. The following data of the ICO Expert Ratings are overall scores, profile scores, team scores, vision scores, and product scores. We removed 30 initial coin offerings (ICOs) that raised an expert evaluation score of zero to improve the efficacy of data analytics. The data collection was compiled using sentiment analysis and ICO expert ratings. To ascertain the relevance and significance of a successful initial coin offering funding by aggregating tweets and expert ratings, the total number of the ICOs to evaluate is decreased to 57 ICO Table 1.

Table 1. Number of ICOs Dataset in Australia and Singapore by Industry Sector

Industry sector	Australia		Singapore	
	Data Sampling	Data set	Data Sampling	Data set
Business service	3	2	14	8
Cryptocurrency	6	3	26	17
Entertainment	2	1	6	5
Platform	7	5	23	16
Total	**18**	**11**	**69**	**46**

Before participating in an ICO, investors should analyze numerous ICO characteristics such as expert ratings, team members behind the project, campaign length, and the proportion accessible for public sale to limit investment risk. Furthermore, investors would be wise to consult specialists who provide ratings and act as information middlemen in the funding of ICO enterprises. Investors should also consider the impact of market sentiment on the success of an ICO and its aftermarket performance. Our findings also give useful information for companies considering launching an ICO. To attract investors and so enhance the possibility of the campaign's success, such enterprises must focus on indicating ICO project quality.

E. Tweet

Social media are tools for creating awareness among investors around the world. ICO issuers use Blockchain technology to create new cryptocurrencies and sell these new digital tokens or rights of the products or services being developed by the ventures to investors in order to support their projects. The information source on social media creates emotions and reactions among investors, which influence their decision to invest in an ICO [20]. The Twitter sentiment is highly correlated with market performance. Both the search trend and an overall number of tweets are indicators of fundraising volume. The relationships between startups' raised funds and Twitter sentiment show increasing emotionality in tweets towards the successful ICO. Twitter is a social networking platform and a popular channel of communication. Sentiment analysis for market prediction may be accomplished by implementing Twitter Tweets [21]. A favorable trend in search terms and an increase in the total number of tweets are both indicators of increased fundraising activity. The success of the ICOs may be determined using data from businesses on social media and ICO rating platforms. The relationship between funds raised by startups and Twitter sentiment shows increasing emotionality in tweets toward the success of the ICO. Twitter is a popular source of data for sentiment analysis, particularly for financial market forecasting. We assume the average positive sentiment of the ICO's tweets is positively linked to the successful ICOs. Scrapy was then used to gather information on the ICO's tweet, which we then used in our python scraping framework. The Twitter data retrieval was conducted by using a keyword as a short name of the ICO to search. Following that, Twitter data for 57 initial coin offerings were to be collected in order to conduct market analysis and forecast the success of fundraising volumes.

F. Emotional dictionary & sentiment

We retrieved market sentiment data from each tweet. Twitter data was obtained and gathered using Twitter APIs in combination with Scrapy and Tweepy. Scrapy is a Python scraping open-source framework and Tweepy can allow filtering based on hashtags to collect relevant data. Twitter data was retrieved using a keyword search for the name ICO. Then, the prefix "$" was added to identify the currency of each token, significantly reducing the number of unnecessary data retrievals. The 57 initial coin offerings generated 68,281 tweets. Each tweet had the following information: (i) ID, (ii) username, (iii) tweet, (iv) URL, (v) number of retweets, (vi) number of favorites, (vii) number of replies, and (viii) date and time. Table 2. shows an example of a collected KNC tweet in a dataset. This can be described as a user ID (ID) 1271800627505459206 whose name (username) as BarlowZach92 tweeted (tweet) that Anybody still trying to play the $ MKR game is playing with fire. I moved to $ KNC which is bullish. This tweet from BarlowZach92 shared the ICO’s about MKR and KNC and possibly persuaded other investors to move to KNC. Moreover, a hashtag dollar symbol can specify the MKR and KNC from other ICOs. The sentiment of the tweet from BarlowZach92 was considered positive. The MKR or Maker Token is the Governance token of the Maker DAO and Maker Protocol, an Ethereum-based lending platform. Kyber Network, also known as KNC, is a Singapore-based blockchain-based liquidity hub that connects liquidity from a variety of sources to power instant and secure crypto exchange in any decentralized application without the need for an intermediary. Hashtags are widely used in tweets regarding an ICO or cryptocurrency, followed by the ticker symbol. Mentions of ICOs can be found on Twitter, and hashtags can also be found on other social media platforms.

Table 2. Example of KNC tweets within a data set

i	ID	1271800627505459206
ii	Username	BarlowZach92
iii	Tweet	Anybody still trying to play the $ MKR game is playing with fire. I moved to $ KNC which is actually bullish
iv	URL	/BarlowZach92/status/1271800627505459206
v	Number of retweets	0
vi	Number of favorites	4
vii	Number of replies	2
viii	Date and time	6/13/2020 20:44:33 PM

Sentiment Classification is one approach for opinion mining. The data from tweets were analyzed in order to classify social media comments using machine learning. The NRC Word-Emotion Association Lexicon is a collection of terms and their connections with two sentiments as negative and positive emotions and eight emotions: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. After retrieving all tweets for each ICO, we will use the R software to determine the emotional polarity of each tweet using the NRC Emotion Lexicon scale. Figure 2. illustrates a bar graph of the KNC ICOs' sentiment ratings in ten categories. This lexicon, which is optimized for Twitter, assigns a real-valued emotion score to each word. The R program's sentiment analysis outputs the scores for various sentiments in Table 3.

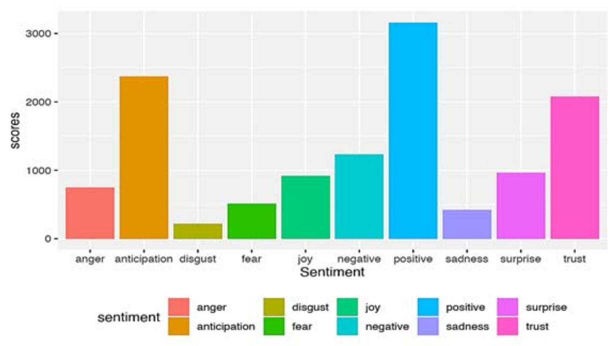

Figure 2. Sentiments Score of Kyber Network (KNC) by R program

Table 3. Example Sentiment Score of Kyber Network (KNC)

Sentiment	Score
anger	749
anticipation	2374
disgust	222
fear	516
joy	923
sadness	423
surprise	964
trust	2081
negative	1232
positive	3161

G. Machine learning methods

Machine learning is classified into two types: supervised learning and unsupervised learning. The supervised learning algorithm requires training data as well as test feedback [22]. To comprehend the link between input and output data. Next, classify future datasets using input training data to forecast the likelihood that subsequent data will fall into one of the established classifications. The algorithm for supervised learning involves training data and test feedback.

To learn the relationship between input data and output data. Next, classify future datasets into categories use input training data to predict the likelihood that subsequent data will fall into one of the predetermined categories. Data interpretation is frequently associated with classification, when a specific object is to be related to one of the previously determined classes, clustering when objects are split into initially undetermined groups (clusters), and forecasting, when it is necessary to predict its future state in space or time using some volume of initial data describing the process background, for example. The machine learning approach employs well-known ML algorithms to solve the SA as a regular text classification problem that employs syntactic and linguistic features. When strict formal classification or clustering techniques are not used, the ML methods are widely used. Beginning with solution trees, genetic algorithms, and metric techniques such as Support Vector Machines (SVM), Logistic Regression (LR), Random Forest (RF), and Naive Bayes (NB), the ML techniques cover a wide range of algorithms. This approach aims to solve the central challenge of an intelligent structure by anticipating all other activities and evaluating the existing object. The existence of labeled training documents is required for supervised learning methods. There are numerous types of supervised classifiers in the literature. In the following subsections, the most commonly used classifiers in the SA are discussed. The dataset was separated into training and testing subsets, which accounted for 70% and 30% of the total, respectively. Machine learning approaches have increased sentiment analysis accuracy and accelerated autonomous data evaluation in recent years. This study sought to use four machine learning approaches for sentiment analysis. The modeling of four strategies is detailed briefly below.

i. Support Vector Machine (SVM)

SVM is a very important machine learning algorithm. This algorithm can do linear and nonlinear classification, regression, and outlier detection [23]. The SVM is another algorithm widely used by machine learning people for both classifications as well as regression problems. It is chosen above other classification algorithms due to its efficiency and accuracy. It delivers trustworthy findings even with limited data. For example, we have two classes to consider: class A: the successful ICO and class B: the unsuccessful ICO. SVM takes into account all data points and generates a line termed the successful ICO' that divides the two groups. The SVM considers all data points and generates a line called the 'successful ICO' that separates the two groups. This is referred to as the 'Decision border'. Anything that comes within the successful ICO class is classified as class A. The SVM separates data classes using hyperplanes. Then it will determine which Hyperplane is the optimal line for data classification (Optimal hyperplane).

$$h_\theta(x) = w_1x_1 + w_2x_2 + \cdots + w_nx_n + b = w^Tx + b \quad (2)$$

ii. Logistic Regression (LR)

LR is a machine learning classification approach. It models the dependent variable using a logistic function. The dependent variable is binary, implying that there are only two potential classifications [24]. By employing the greatest likelihood approach, logistic regression produces the most accurate predictions. Sigmoid is a mathematical function that has the property of mapping any real value between -∞ and + ∞ to a real value between 0 and 1. Thus, if the output of the sigmoid function is more than 0.5, it is classified as positive, and if it is less than 0.5, it is classified as negative. Because of the additive linear combination of the independent variables, the LR is simple to interpret. It is used to divide objects into two classes, such as "negative" and "positive." In this case, the function of the hypothesis necessitates the fulfillment of the condition which is accomplished through the use of a sigmoid (logistic) function.

$$h(z) = \frac{1}{1+exp^{-z}} \quad (3)$$

It maps the input 'z' to a value that ranges between 0 and 1, and so it can be treated as a probability.

iii. Random Forest (RF)

The Random Forest algorithm is a supervised learning model that is versatile [25], even without hyperparameter optimization, this simple machine learning method consistently produces excellent results. A random forest's hyperparameters are very similar to those of a decision tree or a bagging classifier. It's also one of the most widely used algorithms for classification and regression issues. A random forest forecast combines multiple decision trees to produce a more accurate and reliable prediction. The random forest algorithm can be represented mathematically as:

$$Random\ Forest = argmax_{j\in\{1,2,\ldots N\}} \sum_{i=1}^{i} DecisionTrees_{i,j} \quad (4)$$

Where class j is the number of decision trees from 1 to the example and i is the number of classes in the data. The maximum value of the function, or the majority vote, is referred to as Argmax.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(fi - yi)^2 \quad (5)$$

Where N is the number of data points, fi is the value returned by the model, and yi is the actual value for data point i.

iv. Naïve Bayes (NB)

It is well established that Naive Bayes outperforms even the most complex classification techniques [26]. The Naive Bayes classifier assumes that the existence of one feature in a class has no relationship with the presence of any other feature. Naive Bayes classifiers are frequently used in text classification because these perform better in multi-class prediction and have a greater success rate than other algorithms. Based on the distribution of the words in the document, the Naïve Bayes classification model computes the posterior probability of a class. The model employs ICOs feature extraction, which disregards the position of the word in the document. As a result, it is extensively used to identify spam e-mails and to do sentiment analysis on social media to ascertain both positive and negative consumer opinions. Naïve Bayes theorem provides a way of calculating posterior probability $P(c|x)$ from $P(c)$, $P(x)$, and $P(x|c)$ as the equation below:

$$P(c|x) = \frac{P(x|c) \times P(c)}{P(x)} \quad (6)$$

$$P(c|x) = P(x_1|c) \times P(x_2|c) \times \ldots \times P(x_n|c) \times P(c) \quad (7)$$

$$P(c|x) = \frac{P(c)(\prod_{i=1} P(f_i)|c)^{ni(x)}}{P(x)} \quad (8)$$

that f_i are conditionally independent given x class.

3.2. The collection of the dataset

Describe the data source and research methodology. Our ICO dataset was generated using data from the website ico.tokens-economy.com as of June 2020. This website includes the ICO's information: the ICO name, industry categories, fundraising period, website, and total amount raised in the subsequent 29 categories: Art, Artificial Intelligence, Banking, Big Data, Business Services, Casino & Gambling, Charity, Communication, Cryptocurrency, Education, Electronics, Energy, Entertainment, Health, Infrastructure, Internet, Investment, Legal, Manufacturing, Media, Platform, Real estate, Retail, Smart Contract, Software, Sports, Tourism, and Virtual Reality are just a few examples, Figure 3. Between September 2016 and June 2020, we have gathered data in four industries: business services, cryptocurrency, entertainment, and platform. The sampling for 87 of 233 ICOs is included in Table 4.

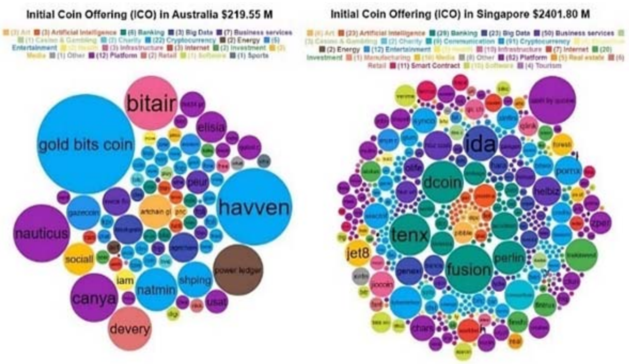

Figure 3. Australia and Singapore ICO data on the website ico.tokens-economy.com

Table 4. Number of ICOs Data Sampling in Australia and Singapore for Industry Sector

Industry Sector	**Australia**		**Singapore**	
	Number of ICO	*Data Sampling*	*Number of ICO*	*Data Sampling*
Business Service	7	3	41	14
Cryptocurrency	19	6	74	26
Entertainment	4	2	12	6
Platform	12	7	64	23
Total	**42**	**18**	**191**	**69**

4. Results and Discussion

Sentiment analysis is a form of natural language processing (NLP) for measuring public mood that automatically recognizes sentiments inside text. Sentiment analysis is performed on Twitter tweets. The insight gained by analyzing unstructured data is vital for product or service developers. Sentiment analysis is used to determine how an online reviewer's attitudes, opinions, and emotions may influence the success of the business [27]. Sentiment analysis helps organizations to better understand their consumers' emotions through social media, customer evaluations, and surveys [28].

Sentiment analysis is the most widely used text classification approach for analyzing incoming data and identifying whether the underlying sentiment is positive, negative, or neutral [29]. Sentiment analysis may help to track trends in the market, and that information can be used to support the growth of businesses and the creation of new products. The use of Twitter to communicate corporate announcements increases investors' attention and reduces information asymmetry [30]. Social media sites like Twitter, Facebook, and YouTube are critical routes for disseminating knowledge into new markets. Big data and natural language processing (NLP) techniques are being used to analyze social media attitudes and to measure the financial performance of initial coin offerings (ICOs) [31].

Hypothesis 1. The average positive sentiment of ICO's tweets is positively linked to the successful ICOs.
Hypothesis 2. The expert ratings informational on the ICO bench website are positively linked to the successful ICOs.
Hypothesis 3. The Sentiment analysis and expert ratings are a Strong positive relationship linked to predict the successful ICOs.

For example, in comparison to review sites or web diaries where the data found isn't only completely useful but moreover requires a number, the proportion of the data existing on Twitter is an unthinkable check. Although 1/60th of all Twitter users may appear to be a little number, it actually amounts to 100 million people. A billion tweets are sent every day, and the number is expanding by the day. This section summarizes the findings of the investigation and examination.

The Orange program is a free and open-source machine learning and data mining toolkit that contains approaches for visualization, exploration, preprocessing, and modeling. The Orange is a simple Python and C++ programming tool for machine learning and data mining. Python modules can help with data mining activities ranging from data preparation to modeling and assessment [32]. Machine learning algorithms are quite effective at classifying tweets. Supervised learning methods give you a strong tool for utilizing machine learning to categorize and analyze data. Classification is a natural language processing activity that uses machine learning techniques to analyze data categories. Classification techniques will be used to analyze data in order to predict. We can do sentiment analysis on tweets by using classification techniques. In this case, classification algorithms were used as a supervised machine learning technique to determine if tweets about the ICOs were favorable, neutral, or negative. Lastly, predictive models were constructed to find an optimized model that can predict the successful ICOs by using the sentiment data from the Twitter and the ICO Ratings. The predictive models were also implemented on the orange program. Among the 57 analyzed ICOs, 34 (59.65%) were successful ICOs and 23 (40.35%) failed ICOs. The business service sector had 6 successful ICOs and 4 failed ICOs. The cryptocurrency sector had 9 successful ICOs and 11 failed ICOs. The entertainment sector had 5 successful ICOs and 1 failed ICOs. The platform sector had 14 successful ICOs and 7 failed ICOs as shown in Figure 4.

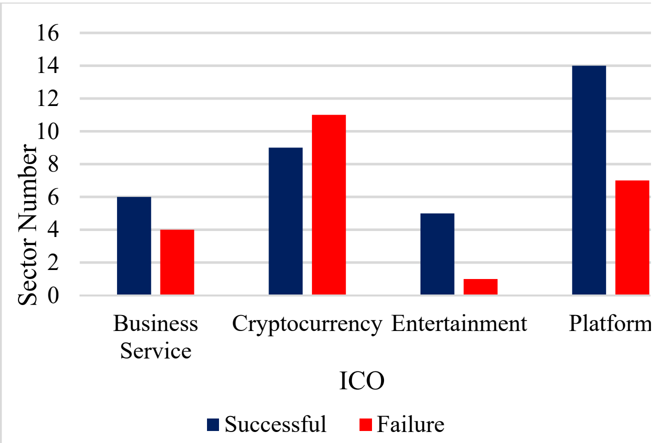

Figure 4. Number of successful and failed ICOs for the Industry sector

As shown in Table 5., Singapore had more successful ICOs than failed ICOs in all industry sectors, while Australia had more successful ICOs than failed ICOs in the Entertainment and Platform sectors. The Total Industry sector in Singapore and Australia had more successful ICOs than failed ICOs in the Business service, Entertainment, and Platform except for the Cryptocurrency sector.

Table 5. Comparison of successful and failed ICOs

Industry sector	Australia		Singapore	
	Success	Failed	Success	Failed
Business service	1	1	5	3
Cryptocurrency	0	3	9	8
Entertainment	1	0	4	1
Platform	5	0	9	7
Total	**7**	**4**	**27**	**19**

Sentiment analysis (AS) results from Twitter data show that positive tweets were associated with the success of the ICOs and referred to fundraising achievement. The predictive models were generated based on four different classification methods consisting of (SVMs), Logistic Regression, Random Forest, and Naïve Bayes. Most likely, the performance improvement is attributable to the data set's domain. While our technique is trained in a cross-validation setting on the data's limit. Cross-validation was used to determine the performance of each model. This study used 5-fold cross-validation, which divides the data into five equal parts. Following that, four parts of the divided data were used to train the predictive model and the remaining sections were used to evaluate the model's performance. The performances of each method are illustrated in the format of confusion matrices as shown in Tables 6. - 9. comprising SVM, Logistic Regression, Random Forest, and Naïve Bayes respectively. As shown in Figure 5., TP (True Positive) and TN (True Negative) refer to the number of correctly detected successful and failed ICOs. The term FP (False Positive) denoted failed ICOs that were incorrectly classified as successful by the model, whereas FN (False Negative) denoted the number of successful ICOs that were incorrectly classified as failed. In order to provide transparent results, we inform all five metrics. Each metric is calculated using the formula that corresponds to it:

$$\text{Accuracy}=\frac{TP+TN}{TP+FP+TN+FN} \quad (8)$$

$$\text{Precision}=\frac{TP}{TP+FP} \quad (9)$$

$$\text{Recall}=\frac{TP}{TP+FN} \quad (10)$$

$$\text{F1}=2\times\frac{Recall\times Precision}{Recall+Precision} \quad (11)$$

Table 6. Confusion Matrix of the SVM

		Actually		
		Positive	*Negative*	∑
Predicted	*Positive*	32	16	48
	Negative	2	7	9
	∑	34	23	57

Table 7. Confusion Matrix of the Logistic Regression

		Actually		
		Positive	*Negative*	∑
Predicted	*Positive*	26	16	42
	Negative	8	7	15
	∑	34	23	57

Table 8. Confusion Matrix of the Random Forest

		Actually		
		Positive	*Negative*	∑
Predicted	*Positive*	27	14	41
	Negative	7	9	16
	∑	34	23	57

Table 9. Confusion Matrix of the Naïve Bayes

		Actually		
		Positive	*Negative*	∑
Predicted	*Positive*	22	10	32
	Negative	12	13	25
	∑	34	23	57

The Performance results of each predictive model are shown in Tables 10. - 12. Table 10. shows the performance sentiment evaluation results of Naïve Bayes have the highest accuracy and recall of 91.2 %, 93.9% precision, and 92.5% F1 score. The SVM has accuracy of 87.7%, 86.5% precision, 94.1 % recall and 90.1% F1 score. The Logistic Regression has lowest accuracy of 87.2%, 80.0% precision, 82.4 % recall and 81.2% F1 score. The Random Forest has accuracy of 85.5%, 87.5% precision, 84.4 % recall and 85.8% F1 score.

Table 10. Performance evaluation results of Sentiment

Model	Accuracy	Precision	Recall	F1
SVM	87.7	86.5	**94.1**	90.1
Logistic Regression	87.2	85.0	82.4	81.2
Random Forest	85.5	87.5	84.4	85.8
Naïve Bayes	**91.2**	**93.9**	91.2	**92.5**

Table 11. shows performance expert rating evaluation results of the SVM has the highest accuracy of 89.5 %, 93.8% precision, 88.2% recall and 90.9% F1 score. The Logistic Regression has accuracy of 82.5%, 85.3% precision, 85.3% recall and 85.3% F1 score. The Random Forest has accuracy of 86.0%, 96.4% precision, 89.4 % recall and 87.1% F1 score. The Naïve Bayes has accuracy of 82.5%, 92.9% precision, 86.5 % recall and 83.9% F1 score.

Table 11. Performance evaluation results of Expert Ratings

Model	Accuracy	Precision	Recall	F1
SVM	**89.5**	93.8	**88.2**	**90.9**
Logistic Regression	82.5	85.3	85.3	85.3
Random Forest	86.0	**96.4**	89.4	87.1
Naïve Bayes	82.5	92.9	86.5	83.9

In Table 12., the SVM was considered to be the high optimal predictive model with Sentiment and Expert Ratings for the successful ICOs with 94.7% accuracy, 94.3% precision, 97.1% recall, and an F1 score as the weighted average of precision and recall at 95.7%. Random Forest achieves the accuracy of 87.7%, 90.9% precision, 88.2% recall, and 89.6% F1 score. Naïve Bayes reaches the accuracy of 91.2%, 91.4% precision, 94.1 % recall, and 92.8% F1 score.

Table 12. Performance evaluation results of Sentiment and Expert Ratings

Model	Accuracy	Precision	Recall	F1
SVM	**94.7**	**94.3**	**97.1**	**95.7**
Logistic Regression	84.2	87.9	85.3	86.6
Random Forest	87.7	90.9	88.2	89.6
Naïve Bayes	91.2	91.4	94.1	92.8

For the parameter of Model evaluation, it is used the Cross-Validation which illustrates the system's ability to create correct new predictions. To measure the efficiency of the models, 5-fold cross-validation was used in the suggested work. The dataset is partitioned into 5 subsets and repeated 5 times in 5-fold cross-validation. Every cycle uses 5 subsets as the training sample and 5-1 subsets as the testing sample. The four models are cross verified. The cross-validation of the three models for 10 runs is presented in Table 12. The accuracy of all four models in all iterations is found to be in the range of ten percent, yielding the accuracy, precision, recall, and F1-score. In Table 13., the SVM achieves high accuracy with Expert Ratings at 93.9%. The comparison of the ICO Success prediction performance is the SVM. The SVM was considered to be the high optimal predictive model with Sentiment and Expert Ratings for the successful ICOs with 93.8% accuracy total. Logistic Regression got the lowest accuracy total at 91.7%. Random Forest realizes an accuracy total of 92.5%. Naïve Bayes have high accuracy with sentiment at 94.2 %.

Table 13. Comparison of the ICO Success Prediction Performance

Model	Accuracy of Sentiment	Accuracy of Expert Ratings	Accuracy Total
SVM	93.7	**93.9**	**93.8**
Logistic Regression	91.2	92.5	91.7
Random Forest	92.9	92.0	92.5
Naïve Bayes	**94.2**	92.5	91.2

For Sentiment and the Expert Rating prediction features are accomplished with high accuracy. We compare four statistical methods to achieve the highest performance of the Naïve Bayes algorithm best result. Results showed close completion between the Naïve Bayes and the SVM at 94.6 (Table 14. - 16.).

Hypothesis 1. The average positive sentiment of the ICO's tweets is positively linked to the successful ICOs. These reputability rating websites are among the first information sources for retail investors looking to distinguish the ICOs that are likely to be scams. However, while these websites provide basic information such as the end date, target volume, and linked social media accounts, more detailed analyses are generally not available.

Table 14. Model Comparison Sentiment for involving machine learning algorithm

Model	SVM	Logistic Regression	Random Forest	Naïve Bayes
SVM	-----	90.8	91.0	94.6
Logistic Regression	85.2	-----	87.7	88.5
Random Forest	90.0	89.5	-----	89.8
Naïve Bayes	87.4	92.5	89.0	-----

Hypothesis 2. The expert ratings information on the ICObench website are positively linked to the successful ICOs. Most ICO investors find the technical information in whitepapers difficult to understand, and most investors lack the time and expertise to conduct their own due diligence on the project. One solution to this problem is to use expert ratings. The ICObench is a well-known cryptocurrency rating website comprised of experts who voluntarily review ICOs. To be considered an expert, one must have a thorough understanding of cryptocurrencies and the underlying market dynamics.

Table 15. Model Comparison Expert Ratings for the use machine learning algorithm

Model	SVM	Logistic Regression	Random Forest	Naïve Bayes
SVM	-----	80.5	86.2	83.1
Logistic Regression	**97.5**	-----	**88.2**	**96.8**
Random Forest	93.8	91.8	-----	92.9
Naïve Bayes	86.9	**93.2**	87.1	-----

Hypothesis 3. The Sentiment analysis and expert ratings are a Strong positive relationship linked to predict success.

Table 16. Model Comparison Sentiment and Expert Ratings for the employ machine learning algorithm

Model	SVM	Logistic Regression	Random Forest	Naïve Bayes
SVM	-----	**98.7**	84.2	82.4
Logistic Regression	81.3	-----	**95.9**	**86.0**
Random Forest	85.8	94.1	-----	84.1
Naïve Bayes	**87.6**	84.0	85.9	-----

The strategy for businesses employing Twitter information should begin in creating storytelling for the followers interested in the story you want to communicate. It will increase people's information on the products, services, and businesses, as well as increase sales opportunities and drive interaction, allowing people to reach out to more businesses more rapidly and follow with market influencers to realize what they think about our company. We can respond promptly to that issue and build a network of people interested in products or services and then increase the number of followers or potential clients in a short time by rewarding followers who take part in the company's marketing activity such as a tweet, retweet, reply to your post and you can monitor brand conversation by Hashtags. As well as being able to discover influencers who have created content and generated follower engagement or gained recognition within the business's target consumers. Finally, using Twitter as a Customer Service Center by helping and resolving issues for consumers who mention our business on Twitter can assist brands in reaching their target audience and impressing their customers. Consequently, the businesses are recommended to consider this factor to make an effective marketing strategy to monitor social comment. Future studies should increase the sample size and time frame of data and focus on evaluating the performance of successful ICOs and expanding the dataset of market sentiment across multiple social media platforms in demand to obtain better accuracy and precision in the results.

5. Conclusion

In this paper, we investigated the success of the ICO's fundraising in the relationship between tweets and expert rating. Positive tweets were related to ICO success and fundraising, according to Twitter sentiment research. Twitter is a prominent social media site for promoting Initial Coin Offerings (ICOs) Most investors are following Twitter for news on investing in ICOs, IEOs, Defi, blockchain technology, cryptocurrency, influencer movements, and crypto transactions. We discovered strong correlations between an ICO's fundraising and its favorable Twitter sentiment and expert rating score on the ICObench website. We discovered that startup businesses with positive comments or more discussions on Twitter had a higher chance of raising more funds. As a result, the benefits linked with social media appear to be related to both an awareness impact, where investors may learn and share information about the ICO or cryptocurrency market. The predictive accuracy of the SVM is determined to be the best among the four classifiers, namely SVM, LR, NB, and RF. The accuracy results were cross-validated, and the greatest amount of accuracy achieved among the four models was 94.7 % for the SVM. In the future, the work might be expanded to do the multiclass classification of reviews, which would provide the consumer with a more demarcated type of review, resulting in a better judgment of the product. It can also be used to forecast a product's rating based on the review. This will supply users with dependable ratings because occasionally the product's rating and the sentiment of the review do not correspond. The suggested work extension will be very useful to the e-commerce industry because it will increase user happiness and trust.

References

[1]. Campino, J., Brochado, A., & Rosa, Á. (2021). Initial Coin Offerings (ICOs): the importance of human capital. *Journal of Business Economics*, 1-38.

[2]. Tiwari, M., Gepp, A., & Kumar, K. (2020). The future of raising finance-a new opportunity to commit fraud: a review of initial coin offering (ICOs) scams. *Crime, Law and Social Change*, *73*(4), 417-441.

[3]. Domínguez, M. D. L. C. C. (2021). Financing of Start-Ups via Initial Coin Offerings and Gender Equality. In *The Fourth Industrial Revolution and Its Impact on Ethics* (pp. 183-197). Springer, Cham.

[4]. Vivion, M., Hennequin, C., Verger, P., & Dubé, E. (2020). Supporting informed decision-making about vaccination: an analysis of two official websites. *Public health*, *178*, 112-119.

[5]. Boreiko, D., & Vidusso, G. (2019). New blockchain intermediaries: do ICO rating websites do their job well?. *The Journal of Alternative Investments*, *21*(4), 67-79.

[6]. Giudici, G., & Adhami, S. (2019). The impact of governance signals on ICO fundraising success. *Journal of Industrial and Business Economics*, *46*(2), 283-312.

[7]. Morkunas, V. J., Paschen, J., & Boon, E. (2019). How blockchain technologies impact your business model. *Business Horizons*, *62*(3), 295-306.

[8]. Daskalakis, N., & Georgitseas, P. (2020). *An Introduction to Cryptocurrencies: The Crypto Market Ecosystem*. Routledge.

[9]. Drobetz, W., Momtaz, P. P., & Schröder, H. (2019). Investor sentiment and initial coin offerings. *The Journal of Alternative Investments*, *21*(4), 41-55.

[10]. 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.

[11]. Choi, G. Y., & Behm-Morawitz, E. (2017). Giving a new makeover to STEAM: Establishing YouTube beauty gurus as digital literacy educators through messages and effects on viewers. *Computers in Human Behavior*, *73*, 80-91.

[12]. Albrecht, S., Lutz, B., & Neumann, D. (2020). The behavior of blockchain ventures on Twitter as a determinant for funding success. *Electronic Markets*, *30*(2), 241-257.

[13]. Mohd-Sulaiman, A. N., & Hingun, M. (2020). Liability risks in shareholders' engagement via electronic communication and social media. *International Journal of Law and Management*, *62*(6), 539-555.

[14]. Appel, G., Grewal, L., Hadi, R., & Stephen, A. T. (2020). The future of social media in marketing. *Journal of the Academy of Marketing Science*, *48*(1), 79-95.

[15]. Calderwood, L. U., & Soshkin, M. (2019). The travel and tourism competitiveness report 2019. World Economic Forum.

[16]. Fisch, C., Masiak, C., Vismara, S., & Block, J. (2021). Motives and profiles of ICO investors. *Journal of Business Research*, *125*, 564-576.

[17]. Prokofieva, M. (2015). Twitter-based dissemination of corporate disclosure and the intervening effects of firms' visibility: Evidence from Australian-listed companies. *Journal of Information Systems*, *29*(2), 107-136.

[18]. Lee, H., Ha, K. C., & Kim, Y. (2018). Marketing expense and financial performance in arts and cultural organizations. *International Journal of Nonprofit and Voluntary Sector Marketing*, *23*(3), e1588.

[19]. Bense, H., Tani, M., Saint-Jean, M., Reyssat, E., Roman, B., & Bico, J. (2020). Elastocapillary adhesion of a soft cap on a rigid sphere. *Soft matter*, *16*(8), 1961-1966.

[20]. Zetzsche, D. A., Buckley, R. P., Arner, D. W., & Föhr, L. (2017). The ICO Gold Rush: It's a scam, it's a bubble, it's a super challenge for regulators. *University of Luxembourg Law Working Paper*, (11), 17-83.

[21]. Hernandez-Suarez, A., Sanchez-Perez, G., Toscano-Medina, K., Martinez-Hernandez, V., Perez-Meana, H., Olivares-Mercado, J., & Sanchez, V. (2018). Social sentiment sensor in twitter for predicting cyber-attacks using ℓ1 regularization. *Sensors*, *18*(5), 1380.

[22]. Wang, Z., Tong, V. J. C., & Chin, H. C. (2014, December). Enhancing machine-learning methods for sentiment classification of web data. In *Asia Information Retrieval Symposium* (pp. 394-405). Springer, Cham.

[23]. Chanaim, S., Srichaikul, W., Rungruang, C., & Sriboonchitta, S. (2019). Forecasting GDP in ASIAN Countries Using Relevant Vector Machines. *Thai Journal of Mathematics*, 201-215.

[24]. Kuchin, Y. I., Mukhamediev, R. I., & Yakunin, K. O. (2020). One method of generating synthetic data to assess the upper limit of machine learning algorithms performance. *Cogent Engineering*, *7*(1), 1718821.

[25]. Dawod, A. Y. (2021). Hand Gesture Recognition Based Sign Language Interpretation in Real-Time. *LINGUISTICA ANTVERPIENSIA*, 2987-3007.

[26]. Arora, M., & Kansal, V. (2019). Character level embedding with deep convolutional neural network for text normalization of unstructured data for Twitter sentiment analysis. *Social Network Analysis and Mining*, *9*(1), 1-14.

[27]. Yu, C. E., & Zhang, X. (2020). The embedded feelings in local gastronomy: a sentiment analysis of online reviews. *Journal of Hospitality and Tourism Technology*, *11*(3), 461-478.

[28]. Jain, P. K., Pamula, R., & Srivastava, G. (2021). A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer Science Review*, *41*, 100413.

[29]. Albarrak, M. S., Elnahass, M., Papagiannidis, S., & Salama, A. (2020). The effect of twitter dissemination on cost of equity: A big data approach. *International Journal of Information Management*, *50*, 1-16.

[30]. Ng, F. K., Wallace, S., Coe, B., Owen, A., Lynch, J., Bonvento, B., ... & McGrath, B. A. (2020). From smartphone to bed-side: exploring the use of social media to disseminate recommendations from the National Tracheostomy Safety Project to front-line clinical staff. *Anaesthesia*, *75*(2), 227-233.

[31]. Fisch, C., & Momtaz, P. P. (2020). Institutional investors and post-ICO performance: an empirical analysis of investor returns in initial coin offerings (ICOs). *Journal of Corporate Finance*, *64*, 101679.

[32]. Park, S., & Woo, J. (2019). Gender classification using sentiment analysis and deep learning in a health web forum. *Applied Sciences*, *9*(6), 1249.