Crypto-currencies narrated on tweets: a sentiment analysis approach
Saeed Rouhani and Ehsan Abedin
Faculty of Management, University of Tehran, Tehran, Iran

Abstract

Purpose – Crypto-currencies, decentralized electronic currencies systems, denote a radical change in financial exchange and economy environment. Consequently, it would be attractive for designers and policy-makers in this area to make out what social media users think about them on Twitter. The purpose of this study is to investigate the social opinions about different kinds of crypto-currencies and tune the best-customized classification technique to categorize the tweets based on sentiments.

Design/methodology/approach – This paper utilized a lexicon-based approach for analyzing the reviews on a wide range of crypto-currencies over Twitter data to measure positive, negative or neutral sentiments; in addition, the end result of sentiments played a training role to train a supervised technique, which can predict the sentiment loading of tweets about the main crypto-currencies.

Findings – The findings further prove that more than 50 per cent of people have positive beliefs about crypto-currencies. Furthermore, this paper confirms that marketers can predict the sentiment of tweets about these crypto-currencies with high accuracy if they use appropriate classification techniques like support vector machine (SVM).

Practical implications – Considering the growing interest in crypto-currencies (Bitcoin, Cardano, Ethereum, Litecoin and Ripple), the findings of this paper have a remarkable value for enterprises in the financial area to obtain the promised benefits of social media analysis at work. In addition, this paper helps crypto-currencies vendors analyze public opinion in social media platforms. In this sense, the current paper strengthens our understanding of what happens in social media for crypto-currencies.

Originality/value – For managers and decision-makers, this paper suggests that the news and campaign for their crypto in Twitter would affect people’s perspectives in a good manner. Because of this fact, the firms, investing in these crypto-currencies, could apply the social media as a magnifier for their promotional activities. The findings steer the market managers to see social media as a predictor tool, which can analyze the market through understanding the opinions of users of Twitter.

Keywords Crypto-currencies, Sentiment analysis, Twitter, Classification

Paper type Research paper

1. Introduction
Cryptocurrencies such as Bitcoin (a decentralized electronic currency system) denote a radical change in financial systems after their creation in 2008 by Satoshi Nakamoto (Nakamoto, 2008). Crypto-currencies stand for an information technology innovation based on the improvement in crypto-graphic protocols and peer-to-peer networks (Ron and Shamir, 2013). Because of their properties, most crypto-currencies are not managed by any government or bank. Like any other currency, their main peculiarity is to facilitate services and goods’ transactions (Grinberg, 2012).

Crypto-currencies are rapidly gaining popularity and attracting a considerable number of users (Matta et al., 2015). In addition, the price and market cap of these assets are getting higher and higher. When these changes and popularities occur, consequently the need to
understand the opinion of the public would be increased. As a result, the aim of this paper is to investigate the general attitude of the public toward this innovation on Web 2.0 services.

Web 2.0 services and social networks like emails, forums, tweets, chats and blogs are broadly used as communication media with convincing results. Sharing knowledge is an essential part of improving as well as learning skills. Using social network platforms, team members find opportunities to obtain more comprehensive information about their peers' expertise (Dessai et al., 2012). Data in social media reflect opinions and ideas regarding every aspect of the globe. It is possible to make deep modifications in people's behaviours in the use of social networks and social media. Twitter is a microblogging service and social network website, which has turned into a vital tool for a vast number of individuals and businesses to share information and communicate with fast growth and adoption. It has further developed into a means to exchange thoughts and notions on investing decisions (Matta et al., 2015).

Investigation of the related works clearly shows that the linkage of sentiment analysis of social media with the concept of crypto-currencies is so limited. As the first effort in 2015, Colianni et al. (2015) proved that Twitter data relating to crypto-currencies can be used to identify market movements. Recently, Degrandea, Karaleviciusa, and Weerdt in the year of 2017 emphasized the utilization of sentiment analysis to predict Bitcoin price. They deployed the lexicon-based sentiment analysis techniques combined with Harvard Psychosocial Dictionary to quantify sentiment (Karalevičius, 2017). Both of the mentioned studies only considered Bitcoin as a crypto-currency and focused on the tweets about Bitcoin and trading.

To the best of our knowledge, there is lack of scientific study to investigate peoples' opinions about the majority of crypto-currencies (e.g. Bitcoin, Cardano, Ethereum, Litecoin and Ripple) and whether the public is welcoming them. Also, there is ambiguity to predict the agree/disagree opinions (negative, neutral and positive) based on people's tweets about different crypto-currencies for future and there is a lack of research that analyses people's opinion on these new currencies. Thus, this paper aims:

- to understand the general sentiment of the public towards five different well-known crypto-currencies; and
- to realize the best classifier to predict the sentiments of tweets.

The current research is designed based on a two-phase methodology and analyses and social media activities or information extracted by the web search media, particularly Twitter to obtain people's opinions on crypto-currencies and then classified them into a number of sentiment groups such as 'positive, negative, or neutral'.

Two different sentiment analysis methods have been applied such as supervised and lexicon-based approach on the tweets of users to automatically analyze people's opinions, sentiments, evaluations and attitudes. To fulfill the above-mentioned gap, this research is designed to answer the following questions:

**RQ1.** What are the social opinions (negative, neutral and positive) about all kinds of crypto-currencies?

**RQ2.** What is the best customized classification technique to classify the tweets based on sentiments?

The layout of the current research is organized as follows: Section 2 briefly reviews the related studies of “Social media,” “Social network analysis,” “Sentiment analysis,” “Twitter” and “Crypto-currencies.” Section 3 describes the research method stages including collecting
data from a well-known social network platform “Twitter” and the data pre-processing methods. In Section 4, analysis and results are displayed. In Section 5, the authors discuss the results. Finally, Section 6 states the conclusions.

2. Background

2.1 Social network analytics

Social network analytics is related to the practice of collecting data from social network platforms and studying the data to support managers, scholars and decision-makers to address particular problems. Social network analytics has been used by a large number of people, such as medical professionals, scientists and business or marketing managers. Automated social network analytics is cost-effective and quick in comparison with old-fashioned media analysis in which data collection and analysis are often done manually (Lee, 2017). The popularity of social network analytics increased when popular social media platforms let firms to access vast volumes of client data from their sites (Lee, 2017).

One traditional aim of social network analytics is to recognize and depict the structure of a network, usually using the graph theory (Wasserman and Faust, 1994). When data are gathered from people in a bounded network, individuals generally find connections with the members of other networks. The presence or absence of such connections is important because these relations define the kind of network such as a friendship network or an information network. Connections display a specific type of relationship between network members. In social network analysis, connections are often directed, wherein if an individual finds alter as a friend, then it signifies the individuals’ out-degree, and when alter recognizes individual as a friend, it represents the individuals’ in-degree (Wasserman and Faust, 1994).

Each connection can also have a weight (Valente, 2010), which could be the strength of the relationship, such as weak or strong, or other information about the link, including the type of advice or information. When connections have weight, it is named “valued network data” (Valente, 2010).

Rather than focusing only on social network structure, scholars collecting network data, generally, present research questions about a wide range of subjects including business, sport, disease and health, politics, financial, marketing, etc. In addition, social network analytics can analyze social network data to obtain customers’ innovative ideas and boost their relationships.

A vast number of open-source tools, commercial toolkits and popular platforms that offer uncomplicated and standard analytics exist for firms. Using these tools, researchers, innovative managers and decision-makers are finding new ways of gathering, merging and analyzing data from social network platforms to understand the business environment, customers and relationships and create new products. In the present technology-driven business environment, enterprises should plan their social network analytics attempts and adjust them frequently. Nonetheless, there exists a lack of typologies that could be used by managers to recognize kinds of analytics and find proper methods needed for analyzing content from social networks (Lee, 2017).

2.2 Sentiment analysis

Sentiment analysis is usually used in a discipline that extracts people’s opinions and feelings from the text data, using natural language processing approaches (Danneman and Heimann, 2014). Unlike conventional data mining methods, sentiment analysis and text mining cope with unstructured data (Oza and Naik, 2016). Furthermore, it is also known as opinion mining, with emphasis on text classification problem. Because of huge number of data extracting feelings, gathering information from web-scale text data can be a highly
costly and challenging task (Fernández-Gavilanes et al., 2016). Initially, there were not much subjective data available on the internet; however, by the advent of social networks in early 2000s, people started to exchange their thoughts through these networks.

Most of the current sentiment analysis approaches mainly focus on classifying the individual tweets as negative or positive. They can be categorized as supervised methods (needing training data) and dictionary-based methods (Saif et al., 2016).

Supervised methods work on the training classifiers including naive Bayes and support vector machines (Saif et al., 2016). They have achieved quite good output; however, obtaining training data could be very challenging (Liu, 2010), especially for the constantly evolving Twitter data. Some ways to address this problem include using the distance supervision method (Go et al., 2009) that makes use of automatically generated training data, where emoticons like “:-)” and “:(” are usually used to label tweets as positive or negative. Nevertheless, automatic labelling of training data may cause errors, having an impact on the performance of classifiers (Speriosu et al., 2011). Another weakness of these methods is their domain dependence, meaning that classifiers trained on data from one domain (e.g. tweets relating to sports) produce unacceptable performance when applied to data from a different domain (e.g. tweets relating to e-commerce) (Aue and Gamon, 2005).

Dictionary-based approaches, also called “lexicon-based methods,” use a pool of pre-coded words to define the text’s semantic alignment (Chiu et al., 2015; Thelwall et al., 2011). According to Sun (2017):

As a main sentiment analysis method, lexicon approach is an unsupervised method, in which the text data are classified into a set of predefined sentiment classes. Sentiment scores of the text are calculated based on a sentiment lexicon, which is a dictionary consisting of words and their corresponding sentiment scores (Sun et al., 2017).

With regards to data modelling, lexicon-based methods are considered as models to define the polarity of the text data. They may not deliver high performance results because the terms in the text data might convey a polarity different than specified in the lexicon (Muhammad et al., 2016). This problem can be avoided through making context particular lexicons to decrease term polarity problem.

Both of the above-mentioned methods are commonly used by scholars. We have also used these popular methods, namely, supervised and lexicon-based approaches, in the present study.

2.3 Twitter

Social networks have become increasingly popular over the years. Some studies focus on the application of sentiment analysis tools in comments and publications posted by the users; they allow everybody to be a potential content producer. Also, Twitter plays a key role as an excellent source of information for this type of work. Hence, Twitter has been becoming more popular among its users, which makes it a great source for analysts. The intentions of users in Twitter are different; for example, some people use it only as a means of discussion to talk about their day-to-day actions; some others use it for professional aims; and finally, others use it to spread nasty content. With the limitation to 280 characters that Twitter imposes to post a tweet, lots of users cannot express their feelings in the best way so they tend to use facial expression. Therefore, with fewer letters, users can express their feelings such as sadness, happiness, shyness, anger and others (Kirilenko and Stepchenkova, 2014; Ruan et al., 2018).

Twitter has the benefit of taking a rich amount of reliable users’ in the moment sentiment and behavior (Capriello et al., 2013; Dodds et al., 2011). In comparison with traditional research
methods like polls and comment cards, Twitter could be a new research model that suffers less from bias related to interactions and recall with human issues (Kirilenko and Stepchenkova, 2014). Asur and Huberman (2010) mention that the collective insight on Twitter could be more precise than other techniques in extracting knowledge like opinion polls and surveys; however, it needs to be analyzed properly in a large enough volume. Additionally, providing approximately real-time access to tweets or comments through the API makes Twitter an appropriate platform for large-scale near real-time sentiment analysis (Wang et al., 2013), which is a popular source among lots of researchers for sentiment analysis (Daniel et al., 2017; Kušen and Strembeck, 2018; Pandey et al., 2017; Xiong et al., 2018).

2.4 Crypto-currencies

Ferguson (2008) states a concise history of the progression of the concept of “wealth.” In the medieval period, wealth was associated to having the means to overcome and plunder so that it was viewed as a result of power; next, with the rise of mercantilism and capitalism, wealth began to be understood as the ownership of material goods like gold or other precious metals and the methods to create production and trade it to obtain more material goods; therefore, money was progressively regarded as the source of power. When the capitalist system was established in the western society, the main symbol of wealth became the possession of money, as its great liquidity permits it to be converted into different kinds of assets. Currently, the hard core of the world’s wealth is focused in financial assets rather than in real assets. In this regard, a successful businessman’s wealth is, generally, assessed based on his or her company’s stock value, instead of his or her vessel and his or her car (Ferguson, 2008; Peng et al., 2018).

The use of crypto-currencies has gained power in response to the observed failures of government and central banks during the 2008 crash (Weber, 2014). Bitcoin and other crypto-currencies may also propose cheaper options to the existing debit and credit card systems (Angel and McCabe, 2015) and whether the growth of crypto-currencies can modify the concept of “wealth” again. Recent studies like Vigna and Casey (2016) claim that the new concepts and technologies that come together with crypto-currencies (like the decentralization of money and the blockchain) have the ability to lead the world into a new economy, which can push the traditional means of economic transactions into the digital realm (Fry and Cheah, 2016; Peng et al., 2018; Vigna and Casey, 2016).

Bitcoin was not the first digital currency; eCash, e-Gold and Flooz were former efforts of a completely virtual approach to make transactions, with e-gold (which was created by Gold and Silver Reserve Inc.) being the most successful among them.

The peer-to-peer electronic monetary system was primarily defined in a concise research paper by Nakamoto (2008), in which the aim of a digital currency is outlined along with how the digital currency could be generated and implemented. Nakamoto (2008) argues the drawbacks of the existing electronic payment system and recognizes the high costs of mediating disputes in the mentioned system. To overcome the intrinsic trust problems regarding the electronic payment system, Nakamoto (2008) claims that a crypto-graphic proof would permit “any two willing parties to transact directly with each other without the need for a trusted third party.” The crypto-graphic proof would offer fraud protection to both sellers and buyers (Blau, 2018). The aim of digital currency was to develop the existing electronic payment system by letting individuals to exchange electronic coin using digital signatures, which perform as the proof of ownership. In January 2009, the first Bitcoin transaction occurred, and more than two years later, several reports expected the circulation of Bitcoin to be more than 6.5m with about 10,000 users (Blau, 2018). Bitcoin is a hybrid of commodity money. Bitcoin is the most popular of what has become known as payment
systems, digital monetary and crypto-currencies that exist online via decentralized, distributed networks that use a shared ledger data technology known as blockchain coupled with secure encryption (Hayes, 2017).

According to coinmarketcap[1], Bitcoin is not the only crypto-currency; more than 100 crypto-currencies have emerged so far, such as Ripple and Ethereum. Therefore, in addition to Bitcoin, we chose four other popular crypto-currencies (Ripple, Ethereum, Cardano and Litecoin) using the market cap as criterion to our research.

3. Research method

In this section, as presented in Figure 1, the authors describe the steps taken to conduct the research. First, the required data were collected from the Twitter website. In the second step, to achieve clean data, five pre-processing phases were conducted. Third, the lexicon-based and supervised approaches were applied to extract the sentiment of each tweet about crypto-currencies. Finally, the obtained results were analyzed. All the above-mentioned steps are described in details in the following sections.

3.1 Data collection

In the present research study, the required data were searched and obtained programmatically by using twitteR[2] package from Twitter (Bose et al., 2017). twitteR package is developed in R programming language and mainly interacts with Twitter web API, which delivers access to the Twitter data. When tweets come in, our system stores them into our database. Therefore, searching for the mentioned crypto-currencies terms (i.e. #Bitcoin, #Cardano, #Ethereum, #Litecoin and #Ripple), we extracted the following numbers of tweets for each of them between December 17, 2017 and January 30, 2018 and saved them in a CSV file (Table I).

3.2 Data pre-processing

As the purpose of our study is to analyze the sentiment and opinion of individuals, retweets of a same person are not considered new opinions and could bias our analysis. Thus, in the pre-processing step, we searched for duplicated entries and removed all of them.

Five pre-processing steps (demonstrated in Figure 2) were conducted as follows:

Figure 1.
Research steps
• **Noise cleaning**: Eliminating meaningless and repetitive words, which are not helping in concluding the result such as RT-@-#-URLs-#-usernames, etc.
• **Tokenizing**: Splitting the text into separate parts, called “Tokens.”
• **Lower-case conversion**: In this step, all the tokens are converted to the lowercases.
• **Remove stop words**: In this step, all the meaningless words are removed.
• **Stemming**: Alternate word suffixes and decrease the words’ length to their root forms such as changing “clean,” “cleaner,” “cleaning” and “cleaned” to the base form of “clean.”

4. Data analysis and results
In this section, first, a lexicon-based approach is applied. Then to evaluate how supervised machine learning can be applied for sentiment analysis of the tweets, we started with assessing which classification algorithm provides a better accuracy.

<table>
<thead>
<tr>
<th></th>
<th>No. of all tweets (including re-tweets)</th>
<th>No. of unique tweets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bitcoin</td>
<td>2,253,155</td>
<td>24,578</td>
</tr>
<tr>
<td>Cardano</td>
<td>90,679</td>
<td>13,689</td>
</tr>
<tr>
<td>Ethereum</td>
<td>909,808</td>
<td>11,731</td>
</tr>
<tr>
<td>Litecoin</td>
<td>330,342</td>
<td>9,853</td>
</tr>
<tr>
<td>Ripple</td>
<td>1,792,297</td>
<td>41,310</td>
</tr>
<tr>
<td>Total</td>
<td>5,376,281</td>
<td>101,161</td>
</tr>
</tbody>
</table>

Table I. Number of collected data

![Pre-processing steps](image-url)
4.1 Lexicon-based sentiment analyses

We analyzed the sentiment score of tweets data set using wordnet[3] which is a package developed in R programming language for the sentiment analysis of English contents. wordnet has been extensively used by former researchers for several different goals (Richardson et al., 1994; Varelas et al., 2005; Reynaud and Safar, 2007; Geum and Park, 2016). It includes a widespread sentiment dictionary and can do sentiment analysis and return the sentiment score for the whole sentence. As a result, one is able to calculate the sentiment score of the tweets according to the score of each word. Each word will have a score associated with it ranging from –1 to +1. A positive score shows a positive sentiment and negative score implies a negative sentiment. Three classes of sentiments and their score range are given in Figure 3.

Accordingly, the tweets are classified into three sentiment groups of “negative, neutral, and positive.”

Analyzing our data set, 11,798 (48 per cent) of the 24,578 Bitcoin tweets were classified as positive, 6,570 (27 per cent) as negative and 6,210 (25 per cent) as neutral.

Of 13,689 Cardano tweets, 7,317 (54 per cent) were classified as positive, 3,723 (27 per cent) as neutral and 2,649 (19 per cent) as negative.

Of 11,731 Ethereum tweets, 5,793 (50 per cent) were classified as positive, 3,327 (28 per cent) as neutral and 2,611 (22 per cent) as negative.

Of 9,853 Litecoin tweets, 5,184 (53 per cent) were classified as positive, 2,407 (24 per cent) as neutral and 2,262 (23 per cent) as negative.

Of 41,310 Ripple tweets, 21,413 (52 per cent) were classified as positive, 10,209 (25 per cent) as neutral and 9,688 (23 per cent) as negative. Table II briefly presents these results.

Table III illustrates a few examples of actual tweets in each sentiment; for shortening the table, all usernames, URLs and hashtags are removed. For example, the tweet of “ripple is literally the worst thing that has happened in my life so far” is classified as a negative. On the contrary, the tweet of “ripple is starting her engine to make a nice ride” is in a positive class, and “Bitcoin and its derivatives use decentralized control as opposed to centralized electronic money/centralized banking systems” is classified as a neutral tweet. Figure 4 shows the breakdown of sentiment analysis results of tweets. The figure obviously demonstrates that positive tweets dominate the sentiment categories. After positive tweets, neutral tweets are the second, followed by negative tweets. Figure 5 presents the proportion of each category. Roughly, we can see that, on average, around 51 per cent of the tweets

Figure 3.
Score range of three classes

<table>
<thead>
<tr>
<th>Sentiment</th>
<th>Negative</th>
<th>Neutral</th>
<th>Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bitcoin</td>
<td>6,570 (27%)</td>
<td>6,210 (25%)</td>
<td>11,798 (48%)</td>
</tr>
<tr>
<td>Cardano</td>
<td>2,649 (19%)</td>
<td>3,723 (27%)</td>
<td>7,317 (54%)</td>
</tr>
<tr>
<td>Ethereum</td>
<td>2,611 (22%)</td>
<td>3,327 (28%)</td>
<td>5,793 (50%)</td>
</tr>
<tr>
<td>Litecoin</td>
<td>2,262 (23%)</td>
<td>2,407 (24%)</td>
<td>5,184 (53%)</td>
</tr>
<tr>
<td>Ripple</td>
<td>9,688 (23%)</td>
<td>10,209 (25%)</td>
<td>21,413 (52%)</td>
</tr>
<tr>
<td>All</td>
<td>23,780 (23%)</td>
<td>25,876 (26%)</td>
<td>51,505 (51%)</td>
</tr>
</tbody>
</table>
express a positive feeling, around 26 per cent have a neutral sentiment and 23 per cent have a negative sentiment on all these crypto-currencies.

4.2 Sentiment analysis by supervised approach
In this section, with the aim of finding the best algorithm to predict the tweets’ sentiment, we used the data sets obtained based on the lexicon-based approach as training data. Then, the classification algorithms (neural networks, decision tree, SVM, naive Bayes and K-NN) were applied for learning the outcomes of the sentiment class of tweets. Then, the results were validated and tested based on their accuracy. Next, the stated data set were divided into train and test (30 per cent for testing and 70 per cent for training) so as to find the best algorithm. It was revealed that SVM was more accurate in predicting the class of new tweets than the other algorithms, with the average accuracy of 82 per cent. Consequently, this algorithm was selected. After SVM, neural networks with the average accuracy of 78 per cent showed slightly better accuracy than naive Bayes with the average accuracy of 75 per cent.

<table>
<thead>
<tr>
<th>Tweets</th>
<th>Sentiment score</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>“ripple is literally the worst thing that has happened in my life so far.”</td>
<td>-0.431</td>
<td>Negative</td>
</tr>
<tr>
<td>“Also I lost $80 on Litecoin so I can be a martyr for the war against crypto”</td>
<td>-0.379</td>
<td>Negative</td>
</tr>
<tr>
<td>“Bitcoin and its derivatives use decentralized control as opposed to centralized electronic money/centralized banking systems.”</td>
<td>0</td>
<td>Neutral</td>
</tr>
<tr>
<td>“How to buy Ripple in India: Guide to purchasing XRP on Indian cryptocurrency market”</td>
<td>0</td>
<td>Neutral</td>
</tr>
<tr>
<td>“ripple is starting her engine to make a nice ride”</td>
<td>0.521</td>
<td>Positive</td>
</tr>
<tr>
<td>“$ADA also known as CARDANO is one of the most exciting projects in 2017. Remember to buy on the dip”</td>
<td>0.5</td>
<td>Positive</td>
</tr>
</tbody>
</table>

Table III.
Example of sentiment evaluation on the tweets

Figure 4.
Tweets analysis results’ graph (lexicon-based)
Finally, K-NN and decision tree showed the lowest accuracy with the average of 46 and 43 per cent, respectively. Table IV displays the results of the classifiers’ validation separately together with the accuracy criterion (percentage of correct predictions or relative number of correctly classified examples). Additionally, the average accuracy of each classifier is presented in Figure 6.

\[
Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100
\]

[True positive (TP), true negative (TN), false positive (FP) and false negative (FN)].

5. Discussion
The objective of this study was to investigate the social opinions about different crypto-currencies and choose the best customized classification technique to categorize the tweets based on sentiment. The results showed that 51 per cent of the people’s beliefs were positive about the five currently used crypto-currencies (Bitcoin, Cardano, Ethereum, Litcoin and Ripple) based on their tweets on the social media. In addition, the negative rate of opinions was 23 per cent on all kinds of crypto-currencies. Also, 26 per cent of the people’s tweets reflected neutral emotion. This shows that they are talking

<table>
<thead>
<tr>
<th>Table IV. Validation of algorithms based on accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Decision tree</td>
</tr>
<tr>
<td>K-NN (K-Nearest Neighbor)</td>
</tr>
<tr>
<td>Naïve Bayes</td>
</tr>
<tr>
<td>Neural networks</td>
</tr>
<tr>
<td>SVM</td>
</tr>
</tbody>
</table>
and writing about crypto-currencies. We further trained the automated method of sentiment analysis to extract the meaning of the tweets’ concept on crypto-currencies. According to the research design, five classification algorithms (decision tree, K-NN, naive Bayes, neural networks and SVM) were examined on the trained set of tweets, which had their sentiment scores from the first stage. The outcomes proved that, averagely, SVM could classify 82 per cent of tweets in their true sentiment class. Therefore, it can be prescribed that the best method for classify new tweets about all kinds of crypto-currencies would be SVM, and the developed approach can understand the sentiment of people’s tweets accurately.

In comparison with the related works on the area of social sentiment analysis in the field of crypto-currencies, this research had two highlighted contributions. First, unlike the previous articles (Colianni et al., 2015; Karalevičius, 2017), which have focused on Bitcion, this research considered a wide range of crypto-currencies literally (Bitcoin, Cardano, Ethereum, Litcoin and Ripple). The second contribution was about putting aside the price attribute and analysis of the tweets in pure mode. The trend of past studies about twitter demonstrates the lens of market analysis and pricing about the subjects (Daniel et al., 2017); however, the current research represents the people’s feelings concerning the new attacks of crypto-currencies to our society.

As there are few studies that analyze the sentiments about crypto-currencies specifically in the context of Twitter social media, our findings could provide several insightful implications for both academicians and practitioners.

Surprisingly, our findings showed that people would talk (tweet) positively about all crypto-currencies. For managers and decision-makers, it suggests that the news and campaign for their crypto in Twitter would affect the people’s perspectives in good manner. Because of this fact, the firms, investing on these crypto-currencies, could apply the social media as a magnifier for their promotional activities.

In terms of prediction, our findings steer the market managers to see social media as a predictor tool, which can analyze the market through understanding the opinions of users of Twitter. In other words, firms with prospector strategy perspective tend more heavily to use this kind of analytical systems to achieve required gains.

![Figure 6. Accuracy of the algorithms for all crypto-currencies](image)
5.1 Implications
To the best of our knowledge, this study is among the first studies, which theoretically argued about the position discovering of crypto-currencies and classification of words about them based on sentiment analysis of Twitter. This study extended the knowledge of crypto-currencies in the financial business profile. Considering the growing interest in crypto-currencies (Bitcoin, Cardano, Ethereum, Litcoin and Ripple), our findings have a remarkable value for enterprises in the financial area to obtain the promised benefits of social media analysis at work. In addition, this study helps crypto-currencies vendors analyze public opinion in the social media platforms. In this sense, the current study strengthens our understanding of what happens in social media for crypto-currencies.

Furthermore, the current study is among the first works to design an approach and tune the best algorithms for the prediction of emotions and opinions through checking people’s tweets. In a similar vein, our study reaches this fact that SVM and neural networks can play role professionally in learning the classification of tweets narrated about wide range of crypto-currencies.

5.2 Limitations
The present research study has a number of limitations as well as opportunities for forthcoming expansion. The first and main limitation deals with the nature of sentiment analysis and text mining itself. Lexicon-based sentiments are calculated by a machine using a dictionary library, so it cannot be precise and meaningful as a judgment by a human. Nonetheless, in the absence of probability to analyze enormous volume of text by human, the analysis of sentiments with machine would be quite useful, as it has been used broadly in the literature (Ahmed et al., 2016; Bhattacharya et al., 2016; Park et al., 2016; Saif et al., 2016). The second limitation, in the present research, is that we focused on English tweets because English enjoys the widespread coverage and is the most commonly used language in the world of science. Even though, we suggest doing a comparable analysis on the tweets in other popular languages (e.g. Spanish or French). Finally, whereas the sentiment analysis implemented for this paper offers valuable visions on the public opinion around different crypto-currencies, there exist opportunities for additional analysis, for example, the clustering of tweets or association rules about the accompaniment of a couple of crypto-currencies.

6. Conclusion
The current research was designed to fulfill the answers of research questions “What are the social opinions (negative, neutral, and positive) about all kinds of crypto-currencies?” and “What is the best customized classification technique to classify the tweets based on sentiment?” The research findings proved that more than half of the people’s beliefs are positive about crypto-currencies. Furthermore, this research confirmed that the marketers could predict the sentiment of tweets about these crypto-currencies with high accuracy if they would use the appropriate classification technique. Highlighted achieved contributions are:

- demonstrating and clarification of tweets about a wide range of crypto-currencies;
- offering the approach for sentiment analysis of the people’s words of mouth about crypto-currencies; and
- adopting the best technique for sentiment exploration of people’s talks about crypto-currencies by high accuracy.
We believe that room still remains for future validation and improvement. Thus, further research in other contexts and social media is necessary to fine tune the findings. Based on the limitations, which were discussed before, we suggest:

- applying more supervised and non-supervised classification and clustering on the tweets using text mining tools;
- exploring the association rules about accompaniment of crypto-currencies with each other; and
- including English, Spanish and French tweets to be able to compare and contrast the tweets from different communities and cross-analyze them, in future research works.

We hope that the applied approach and the findings of this paper would be beneficial for both academics and professionals.

Notes
1. https://coinmarketcap.com
2. https://cran.r-project.org/package=twitteR
3. https://cran.r-project.org/web/packages/wordnet

References


Corresponding author
Saeed Rouhani can be contacted at: SRouhani@ut.ac.ir

For instructions on how to order reprints of this article, please visit our website: www.emeraldl集团旗下publishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com