

Aggregating Users' Online Opinions Attributes and News Influence for Cryptocurrencies Reputation Generation

Achraf Boumhidi

(Computer Science Department, LISAC Laboratory, Faculty of Sciences Dhar EL Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco

 <https://orcid.org/0000-0002-7396-9651>, achraf.boumhidi@usmba.ac.ma)

Abdessamad Benlahbib

(Computer Science Department, LISAC Laboratory, Faculty of Sciences Dhar EL Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco

 <https://orcid.org/0000-0002-0039-7832>, abdessamad.benlahbib@usmba.ac.ma)

El Habib Nfaoui

(Computer Science Department, LISAC Laboratory, Faculty of Sciences Dhar EL Mahraz, Sidi Mohamed Ben Abdellah University, Fez, Morocco

 <https://orcid.org/0000-0002-5816-0897>, elhabib.nfaoui@usmba.ac.ma)

Abstract: Reputation generation systems are decision-making tools used in different domains including e-commerce, tourism, social media events, etc. Such systems generate a numerical reputation score by analyzing and mining massive amounts of various types of user data, including textual opinions, social interactions, shared images, etc. Over the past few years, users have been sharing millions of tweets related to cryptocurrencies. Yet, no system in the literature was designed to handle the unique features of this domain with the goal of automatically generating reputation and supporting investors' and users' decision-making. Therefore, we propose the first financially oriented reputation system that generates a single numerical value from user-generated content on Twitter toward cryptocurrencies. The system processes the textual opinions by applying a sentiment polarity extractor based on the fine-tuned auto-regressive language model named XLNet. Also, the system proposes a technique to enhance sentiment identification by detecting sarcastic opinions through examining the contrast of sentiment between the textual content, images, and emojis. Furthermore, other features are considered, such as the popularity of the opinions based on the social network interactions (likes and shares), the intensity of the entity's demand within the opinions, and news influence on the entity. A survey experiment has been conducted by gathering numerical scores from 827 Twitter users interested in cryptocurrencies. Each selected user assigns 3 numerical assessment scores toward three cryptocurrencies. The average of those scores is considered ground truth. The experiment results show the efficacy of our model in generating a reliable numerical reputation value compared with the ground truth, which proves that the proposed system may be applied in practice as a trusted decision-making tool.

Keywords: Reputation generation, Decision-making system, Opinion mining, Sentiment analysis, Cryptocurrency

Categories: I.2, I.7, H.4, M.7

DOI: 10.3897/jucs.85610

1 Introduction

Cryptocurrencies are the native assets of a blockchain network that can be traded, used as a store of value, and utilized as a medium of exchange [Abu-Bakar 2018]. They are issued by the blockchain protocol and they are often referred to as a blockchain's native cryptocurrency. Cryptocurrencies are typically decentralized [Croman et al. 2016], which means that they are not relying on a central issuing authority, instead, cryptocurrencies rely on code to manage transactions. The cryptocurrency market has evolved erratically and at remarkable speed over the course of its short lifespan. As of 2021, more than 6,000 cryptocurrencies were developed and utilized by millions of users¹. People tend to go to Twitter to share opinions and reviews about specific cryptocurrencies and also to consume and view other people's opinions for a better investment decision. This large amount of shared opinions are considered valuable for the purpose of generating the reputation of those cryptocurrencies.

Previous reputation systems have mainly focused on generating the reputation of products and services such as movies, hotels, and restaurants [Abdel-Hafez et al. 2012, Yan et al. 2017, Benlahbib and Nfaoui 2019, Elmurngi and Gherbi 2020, Benlahbib and Nfaoui 2020, Boumhidi and Nfaoui 2021, Boumhidi et al. 2022, Boumhidi et al. 2021], yet they neglected the important domain of cryptocurrencies investment where millions of people share online opinions. A statistical study conducted by Forbes² shows that 17% of Americans get their news from Twitter. Also, a Twitter study³ of 3.7 million Twitter accounts showed that tweets with finance content gained 55% more engagement than those without. In this paper, we proposed the first system in the literature that generates the reputation of cryptocurrencies from opinions and reviews shared on Twitter. Every opinion shared by a user on Twitter consists of various features such as the textual content of the tweet expressed in informal language, images associated with the tweet, and the number of likes and shares received for the tweet. Employing all those features is essential to generate a credible reputation value for a cryptocurrency. Moreover, most people tend to use sarcasm when expressing their online opinions, which could affect the accuracy and reliability of the reputation generation system. Therefore, dealing with sarcasm should be an essential part of a reliable reputation system. Furthermore, other important features should be considered in the domain of digital finance, such as the influence of news and events on the target entity, as well as the volume of demand for that specific entity which indicates the amount of users' interest. In this paper, the proposed reputation system is able to generate a numerical value between 0 and 10, that reflects the reputation of a target cryptocurrency by employing all the aforementioned features. First, we extract the sentiment orientation of the textual content of the tweet using a generalized auto-regressive pretraining for language understanding named (XLNet) [Yang et al. 2019], then we propose an approach for detecting sarcasm opinions based on the sentiment of the text, image, and emojis associated with the tweet. Next, we calculate a popularity score based on the likes and shares received for the tweet. Finally, we combine the previous outputs with a computed cryptocurrency demand score and a news influence score to generate a reliable and trustworthy numerical reputation value. This paper is organized as follows. Section II presents the related work concerning the previous reputation generation systems as well as the sentiment analysis models. Section

¹ <https://www.statista.com/statistics/863917/number-crypto-coins-tokens/>

² <https://www.forbes.com/sites/petersuciu/2019/10/11/more-americans-are-getting-their-news-from-social-media/?sh=61c5470a3e17>

³ <https://blog.hootsuite.com/twitter-statistics/>

III presents the preliminaries. Section IV describes our proposal. Section V details the experiments. And finally, Section VI concludes this paper.

2 Related Work

This section presents the literature review of reputation generation systems and the different techniques used in this field throughout the years. It also addresses the related work of document-level sentiment analysis since it is a key component of reputation generation systems based on opinion mining.

2.1 Reputation Generation Systems based on Text Mining

The reputation of an entity is the product of aggregated experiences of a group of individuals. It is based on past interactions and observations of that entity by the consumers. Reputation generation systems tend to compute a reputation value from user-generated content expressed online. The first form of reputation generation system was presented in [Resnick and Zeckhauser 2002], where a numerical score is computed by aggregating the number of positive ratings and negative ratings separately and keeping a total score as the positive score minus the negative score. This technique was used by the eBay website. A slightly advanced scheme was proposed in [Schneider et al. 2000] which computes a weighted average of all the ratings, where the rating weight can be determined by factors such as the age of the rating, the distance between the rating and current score, etc. In the past few years, researchers developed several types of reputation systems, where they exploited other features such as the textual opinions of the users. Authors in [Abdel-Hafez et al. 2012] proposed the first system that generates a reputation value of products and their features based on users' textual opinions rather than users' ratings. Authors employ opinion mining techniques to extract the sentiment orientation of the opinion as well as the strength of the opinion. Those features are used to compute the reputation value of the products. In [Yan et al. 2017], the authors proposed a system that combines opinion fusion and semantic analysis to generate the reputation of Amazon's products. Those opinions are grouped into several fused principal opinion sets that contain opinions with a similar or the same attitude or preference by using Latent Semantic Analysis (LSA) model and cosine similarity. Finally, they normalize the reputation of the target entity by aggregating the ratings attached to the fused opinions. In [Benlahbib and Nfaoui 2019], the authors proposed an improved version of the work introduced in [Yan et al. 2017] where textual opinions are separated into positive and negative based on their sentiment polarity by applying the two classifiers Naïve Bayes and Linear Support Vector Machine (LSVM). Then the positive and negative reviews are grouped into principal opinion sets based on their semantic relations. Next, they calculated a custom reputation value separately for positive and negative groups to finally compute the final reputation value using the weighted arithmetic mean. In [Benlahbib et al. 2019], the authors proposed a reputation system that separates reviews collected from E-commerce websites into two groups: positive and negative based on their sentiment orientation using the Logistic Regression classifier. Then, they calculated the reputation value based on the statistics of each group. In [Elmurngi and Gherbi 2020], the authors proposed a system that computes reputation scores from users' feedback based on a sentiment analysis model (SAM). The reputation score of a product is the ratio of the number of positive reviews over the total number of reviews toward this product. Another reputation generation system was proposed in [Boumhidi et al. 2021] where they converted the numerical ratings given by

a user on some review websites into textual words based on the intensity of the rating. Then, they fused the output with the textual user's review. Next, they extracted the sentiment polarity of the fused movie reviews using bidirectional long short-term memory (Bi-LSTM) classifier. Finally, they computed the reputation score based on the classification results. In [Benlahbib and Nfaoui 2020], the authors proposed a refined reputation system that generates reputation toward various entities (products, movies, TV shows, hotels, restaurants, and services) by mining customer reviews expressed in e-commerce websites. The system incorporates four review features: review helpfulness, review time, review sentiment polarity, and review rating. First, they computed review helpfulness and review time scores, and they fine-tuned a Bidirectional Encoder Representations from Transformers (BERT) model to predict the review sentiment orientation. Next, they designed a formula to assign a numerical score to each review. Finally, they proposed a new formula to compute reputation value toward the target entity. Experimental results using several real-world datasets of different domains collected from IMDb, TripAdvisor, and Amazon websites show the efficacy of the proposed method in supporting the customer decision-making process compared to state-of-the-art reputation systems. In [Gupta et al. 2020], the authors proposed a system that generates the reputation for books by extracting the sentiment polarity of textual reviews using three models (BERT, Naïve Bayes, and SVM) for more accurate sentiment prediction. Authors in [Boumhidi and Nfaoui 2020] proposed a system that generates reputation from reviews collected from Twitter by extracting their sentiment orientation using the deep learning Bidirectional Encoder Representations from Transformers (BERT) as an embedding layer and a multi-layer Gated recurrent units (GRU) which learns from the representations produced by the transformer. The authors also exploited the emojis expressed within the reviews as a feature to improve the accuracy of extracting reviews' sentiment, which lead to the computation of a trusted reputation value. Experimental results conducted on two Twitter datasets show that the proposed system provides accurate and reliable reputation values. Authors in [Sejung and Han Woo. 2021] provided visuals and semantic network analysis in order to determine strategic ways of communicating coins on social media adopted by popular cryptocurrency companies. They comparatively investigated the sentiment flow over time and language usage patterns between companies with high reputation and firms with low reputation. In addition, they explored the relationship between reputation and marketing communication strategies in terms of emotional intensity of expressions and language usage patterns. Nevertheless, no approach was proposed to compute numerical reputation values toward cryptocurrencies. In [Benlahbib and Nfaoui 2021], the authors proposed a reputation system dedicated to generate the reputation of movies and TV shows by employing fine-grained sentiment analysis where reviews are classified into five classes: strongly negative, weakly negative, neutral, weakly positive, and strongly positive. Then, ELMO (Embeddings from Language Model) and cosine similarity were applied to extract the semantic similarity between reviews and to compute a custom score for each emotion class. Finally, the Weighted Arithmetic Mean is used to compute the movie or TV Show's reputation value. Experimental studies showed that the proposed system outperforms the reputation system proposed in [Yan et al. 2017]. Authors in [Boumhidi and Nfaoui 2021] proposed a system that generates a reputation value toward different entities from user-generated data expressed on the Twitter microblogging platform. The proposed system incorporates the sentiment orientation of the textual reviews, the sentiment intensity of the positive reviews, and the popularity of the users and the tweets. The output is a numerical value between 0 and 10 that reflects the reputation of a specific entity. The authors compared the proposed system's output with weighted average votes taken from IMDb, Amazon, TripAdvisor, and Yelp concerning respectively

four products and services. The comparison shows that the proposed system produces an accurate reputation value that could be used in real-life applications. Recently, the authors in [Boumhidi et al. 2022] developed a reputation system that incorporates spam filtering, review popularity, review posting time, and aspect-based sentiment analysis to generate a numerical reputation value toward products and services. Their proposed system is capable of generating numerical reputation values for an entity and its aspects based on opinions collected from various platforms. Experiment results conducted on multiple datasets show the efficacy of their system compared with the state-of-the-art reputation systems. [Benlahbib et al. 2022] surveyed reputation systems published between 2004 and 2021. They investigated the different NLP techniques applied and the features exploited in each reputation system. Besides, they discussed the limitations and drawbacks of each work. Finally, they presented many areas that still need further investigation and solutions. As far as we know, this is the first attempt to generate reputation toward cryptocurrencies based on the aggregation of various features extracted from Tweets. Table 1 summarizes the opinion mining techniques exploited during the reputation generation and visualization process for the previous reputation systems and for this study.

Work	Language	Domain	Semantic analysis	Sentiment analysis
[Abdel-Hafez et al. 2012]	N/A	N/A	N/A	N/A
[Yan et al. 2017]	- English - Chinese	Products	Latent Semantic Analysis (LSA)	N/A
[Benlahbib and Nfaoui 2019]	English	Movies	Latent Semantic Analysis (LSA)	- Naïve Bayes - Linear Support Vector Machine
[Elmurngi and Gherbi 2020]	English	Products	N/A	Logistic Regression
[Benlahbib and Nfaoui 2020]	English	- Products - Movies & TV Shows - Hotels	N/A	Bidirectional Encoder Representations from Transformers (BERT)
[Gupta et al. 2020]	English	- Movies - Books	N/A	- Bidirectional Encoder Representations from Transformers (BERT) - Naïve Bayes - Support Vector Machine
[Benlahbib and Nfaoui 2021]	English	Movies & TV Shows	Embeddings from Language Models (ELMo) ⁴	Multinomial Naïve Bayes
[Boumhidi and Nfaoui 2021]	English	- Products - Services - Hotels - Movies	N/A	Bidirectional Encoder Representations from Transformers (BERT)
[Boumhidi et al. 2022]	English	- Products - Services - Hotels - Movies	N/A	LCF-ATEPC ⁵
This study	English	Cryptocurrencies	N/A	XLNet

Table 1: Summary of reputation systems based on NLP techniques

2.2 Sentiment Analysis on Stanford Sentiment Treebank dataset

Sentiment analysis has been a hot research topic in natural language processing and data mining fields in the last decade. It is a major task of natural language processing where attitude, opinion, or feeling toward a specific entity is extracted. Sentiment analysis can be characterized into three levels: document-level, sentence-level, and aspect-level [Behdenna et al. 2018]. Since our proposed approach is based on the extraction of sentiment by finetuning XLNet on the Stanford Sentiment Treebank (SST) dataset. This section aims at providing the significant contributions recently made in the domain of document-level sentiment analysis applied to the SST dataset. Generally speaking, document-level sentiment analysis aims to determine the sentiment polarity or intensity of a document by extracting document features and capturing sentence semantic relationships. In [Vaswani et al. 2017], Google researchers revolutionized the field of sentiment analysis by proposing a deep learning model named Transformers that adopts the mechanism of self-attention, which later inspired Jacob Devlin and his colleagues to develop a Bidirectional Encoder Representations from Transformers (BERT) [Devlin et al. 2018] model that was considered a building block for future sentiment analysis approaches. In [Devlin et al. 2018], authors finetuned the BERT-large model by adding an additional output layer to create state-of-the-art models for a wide range of tasks including sentiment analysis in which it achieved an accuracy of 94.9% on the SST-2 binary classification dataset. In [Liu et al. 2019], researchers extend a Multi-Task Deep Neural Network (MT-DNN) originally proposed in [Liu et al. 2015] by incorporating BERT as its shared text encoding layers. The model consists of text encoding layers that are shared across all tasks, while the top layers are task-specific that combine different types of natural language understanding tasks such as pairwise text classification, text similarity, and relevance ranking. MT-DNN obtains a new state-of-the-art result on the SST-2 dataset where it achieved an accuracy of 95.6%. In [Raffel et al. 2019], Google researchers proposed Text-to-Text Transfer Transformer (T5). The model was pre-trained on the Common Crawl's web crawl corpus (C4) and it achieved state-of-the-art results on many natural language processing tasks while being flexible enough to be fine-tuned to a variety of important downstream tasks such as sentiment classification. The T5-large model achieved an accuracy of 96.3% on the SST-2 binary classification dataset. In 2018, a model also built on BERT named Roberta was proposed in [Liu et al. 2019] where it modifies key hyperparameters by removing the next-sentence prediction objective and by training with much larger mini-batches and learning rates. The authors find out that hyperparameter choices have a significant impact on the final results. The model achieves state-of-the-art results on GLUE [Wang et al. 2018] and SQuAD [Rajpurkar et al. 2016]. Recently a new model was proposed in [Clark et al. 2020] named ELEC-TRA which is a pretraining approach that trains two transformer models: the generator and the discriminator. The generator's role is to replace tokens in a sequence and is therefore trained as a masked language model. The discriminator is the model that tries to identify which tokens were replaced by the generator in the sequence. As a result, the contextual representations learned by this approach outperform the ones learned by BERT using the same model size and data. The model also achieved an accuracy of 96.9% on the SST-2 dataset. In [Yang et al. 2019], a new model named XLNet was proposed which is an extension of the Transformer-XL [Dai et al. 2019] model pre-trained using an auto-regressive method to learn bidirectional contexts by maximizing the expected likelihood over all permutations of the input sequence factorization order. XLNet outperforms BERT on 20 tasks, often by a large margin, including sentiment analysis where it achieved an accuracy of 97% on the SST-2 dataset. In [Lan et al. 2019],

researchers proposed a model named ALBERT that employs two parameter-reduction techniques to reduce memory consumption and increase the training speed of BERT by splitting the embedding matrix into two smaller matrices and using repeating layers split among groups. As result, their best model establishes new state-of-the-art results on different benchmarks while having fewer parameters compared to BERT-large.

3 Problem statement

The goal of this work is to compute a numerical value that will represent the reputation of an entity (cryptocurrency) based on user-generated data collected from Twitter, as well as numerical and statistical data related to the target entity. The set of textual content of the tweets toward cryptocurrency j denoted $R_j = \{r_1, r_2, r_3, \dots, r_i\}$ is collected, processed, and fed to an XLNet-based pre-trained sentiment classifier. Next, a rule-based sarcasm detection technique is employed where the set of images associated with each textual opinion in the R_j set is fed to a ResNet50-based image sentiment classifier. The results of both text and image classification including the emojis' sentiment orientation associated with the textual opinion lead to the identification of the authentic sentiment of the textual opinions. Based on the results of the sentiment identification phase, a sentiment score SS_j is calculated. Further, a popularity score PS_j is computed based on the set of likes $L_{ij} = \{l_{1j}, l_{2j}, \dots, l_{nj}\}$ and the set of opinion shares $S_{ij} = \{s_{1j}, s_{2j}, \dots, s_{nj}\}$ received for the tweets. Additionally, a cryptocurrency demand score DS_j is computed based on the result of the entity's demand classification on the R_j set. Lastly, a news influence score NS_j is calculated from the collected rate of change set $ROC_j = \{roc_{d1j}, roc_{d2j}, roc_{d3j}, \dots, roc_{d_{nj}}\}$. The proposed reputation system incorporates the sentiment score SS_j , the popularity score PS_j , the cryptocurrency demand score DS_j , and the news influence score NS_j to generate a final reputation value $Rep_j \in [0, 10]$.

4 Proposed Approach

4.1 System Overview

Figure 1 illustrates the pipeline of our reputation generation system. The proposed system consists of several components and steps, starting by collecting a dataset of opinions (tweets) from Twitter related to a specific cryptocurrency. The dataset is then cleaned and processed, and ready for further analysis. The first step of the proposed system is to compute a numerical sentiment score for the collected opinions based on the results of sentiment classification performed by a fine-tuned auto-regressive language model (XLNet) on the SST dataset, combined with a rule-based sarcasm detection technique. Simultaneously, a cryptocurrency demand score is calculated based on the cryptocurrency demand classification results. Then, a news influence score of the target cryptocurrency is computed based on the historical pricing data collected from a well-known historical price tracker website named CoinMarketCap ⁶. Finally, a popularity score is calculated by incorporating the tweet's social interaction statistics: number of likes (favorites) and number of shares (retweets). Our system aggregates the calculated scores and generates a reputation value for the target cryptocurrency.

⁶ <https://coinmarketcap.com/>

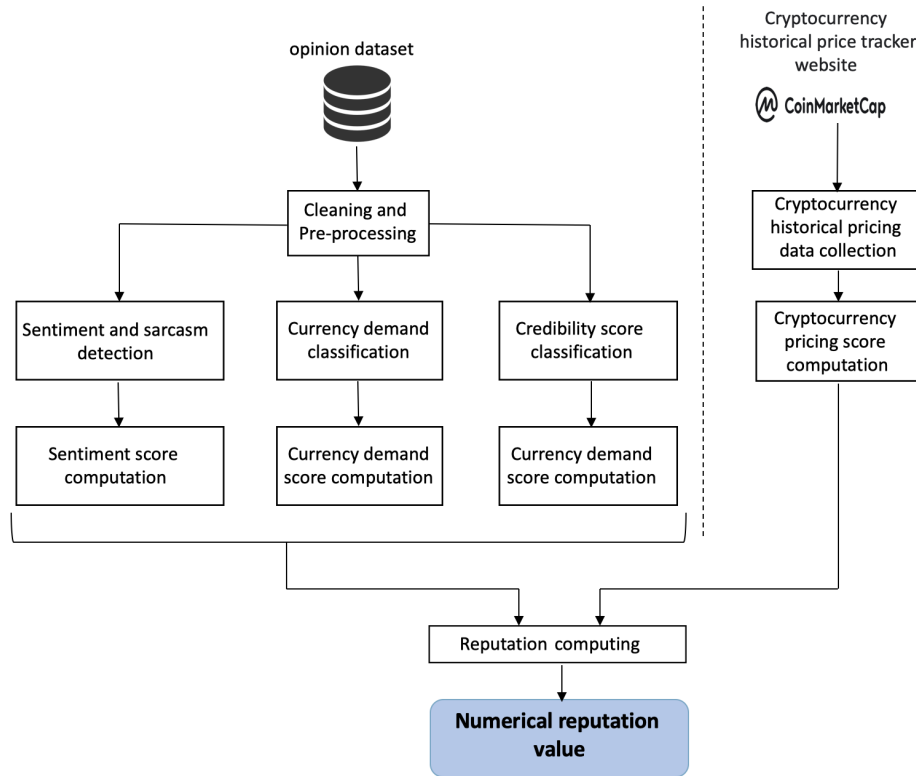


Figure 1: Proposed System Pipeline

4.2 Data collection and processing

Every opinion (tweet) collected from Twitter consists of several components including the textual content of the tweet, images associated with the tweet, and emojis within the textual content. Additionally, users' interactions with the opinions such as the number of likes and the number of shares are also collected. Figure 2 shows a Twitter sample of a collected opinion with its features. The textual reviews are automatically cleaned by filtering unwanted outliers, removing links, eliminating special characters, and replacing slang words. statistical details about the data collection and cleaning are described in section 5.

4.3 Sentiment classification and sarcasm detection

The purpose of employing a sentiment classifier in the proposed system is to extract the sentiment polarity (positive, negative, neutral) of each textual opinion collected from Twitter. This later will go through the process of computing a sentiment score that will be used to calculate the final reputation value. We used a fine-tuned auto-regressive language model named XLNet, which is an unsupervised language representation learning method based on a new generalized permutation language modeling objective. In this paper,

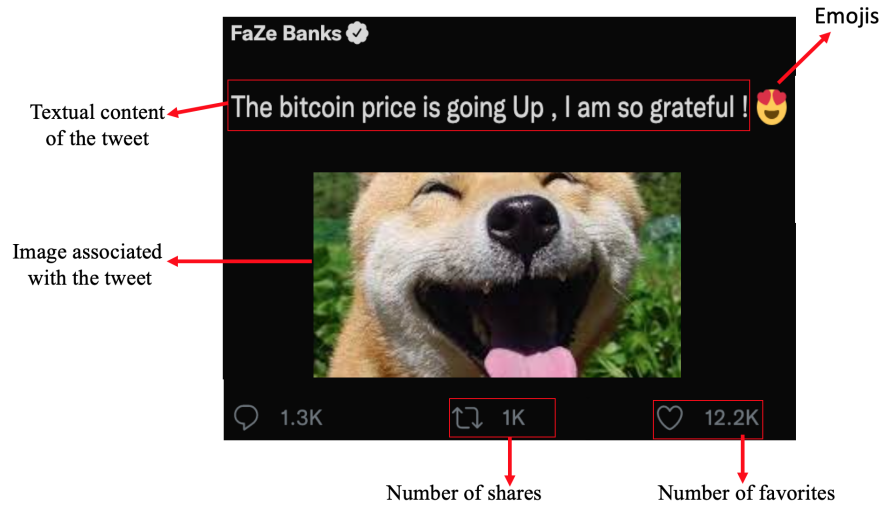


Figure 2: Sample of a collected opinion from Twitter

we finetuned the XLNet model on the well-known Stanford Sentiment Treebank SST dataset, where it achieved state-of-the-art results compared with the previous models. Sarcasm is a complex form of irony used in micro-blogging websites specially on Twitter. It is used to express a negative sentiment disguised as a positive form of multimedia content (text, images, emojis, etc.) and vice versa. Figure 3 shows an example of a negative sarcastic tweet. Indeed, while the textual content of this tweet holds a positive sentiment, the image and the emojis reflect a negative sentiment toward the “Ethereum” cryptocurrency. Therefore, disregarding the detection of sarcasm in the opinion dataset leads to deceitful sentiment orientation of the opinions, which causes the proposed system to generate unreliable reputation values. Therefore, three features of the tweet were exploited to detect sarcasm in opinions: sentiment orientation of the text, sentiment orientation of the image associated with the text, and finally, the emojis contained in the text.

Since the sentiment orientation of the textual content of the tweet was already extracted using the XLNet classifier as discussed above, each associated image will be classified into three possible categories: positive, negative, and neutral by employing a finetuned residual neural network ResNet50 [He et al. 2016] which is a convolutional neural network that is 50 layers deep, used as a backbone for many computer vision tasks especially image classification. ResNet50 is a variant of the ResNet model which has 48 Convolution layers along with 1 Max-Pool and 1 Average-Pool layer. To make the architecture of ResNet50 suitable for fine-tuning, we added a new fully-connected layer head to ResNet, then we trained the model on our created dataset containing 1452 images usually used in sarcastic tweets. This dataset was manually collected and labeled. The emojis are also classified into positive, negative, or neutral using the Emosent python library⁷. Figure 4 summarizes the previous procedures. Finally, for detecting sarcasm, we applied the following rules described in Table 2, where each tweet is classified as sarcastic or not sarcastic. For example, if the text sentiment is positive, the image sentiment

⁷ <https://pypi.org/project/emosent-py/>



Figure 3: Example of a negative sarcastic tweet

is negative, and the emojis sentiment is negative then rule 1 is satisfied, and the tweet is considered sarcastic with an assigned negative label as sentiment polarity.

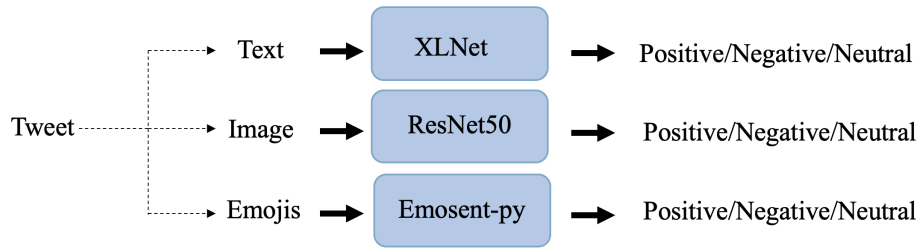


Figure 4: Summarization of sentiment and sarcasm detection procedure

In case the tweet had no associated image or the sentiment of the image is neutral, its sentiment orientation will be judged based on only the result of the textual sentiment prediction, and its emojis will be dismissed. However, in case a tweet has only the textual content and the associated image without the emojis, then rule 7 and 8 will be applied to detect sarcasm. Also, in a case where there is more than one emoji, only the first emoji is considered.

The next step is to calculate a sentiment score based on the total number of positive and negative tweets, which will be used to compute the final reputation value of the target entity. We proposed Eq 1. to calculate the sentiment score by dividing the value of positive tweets P_j by the total value of positive and negative tweets T_j , and by multiplying it by 8.5 in order to bound the output between 0 and 8.5. The max bound value of 8.5 represents the maximum value for the sentiment score in the final reputation generation formula that will be discussed later in section 4.8. In fact, we consider the

	Text sentiment	Image sentiment	Emojis sentiment	sarcasm classification	Overall tweet's sentiment
Rule 1	positive	negative	negative	sarcastic	negative
Rule 2	negative	positive	positive	sarcastic	positive
Rule 3	positive	positive	negative	not sarcastic	positive
Rule 4	negative	negative	positive	not sarcastic	negative
Rule 5	positive	negative	positive	not sarcastic	positive
Rule 6	negative	positive	negative	not sarcastic	negative
Rule 7	negative	positive	-	sarcastic	positive
Rule 8	positive	negative	-	sarcastic	negative

Table 2: Rules employed to detect sarcasm within the tweets

sentiment score more important than the other scores (demand score, popularity score, news influence score).

$$SS_j = \frac{P_j \times 8.5}{T_j} \quad (1)$$

Where:

SS_j : Sentiment score toward entity j .

P_j : Total number of positive tweets toward entity j .

T_j : Total number of tweets toward entity j .

4.4 Cryptocurrency demand score

Unlike traditional currencies, cryptocurrencies are not issued by central banks or backed by the government; therefore, the inflation rates, monetary policy, and economic growth measurements that typically influence the value of currency do not apply to cryptocurrencies. In fact, one of the main aspects that determine the cryptocurrency's value is its demand by individuals. If the demand for an asset increases faster than the supply, the price goes up⁸ which leads to an increase in value, and the currency ends up with a favorable reputation. Therefore, the goal of this sub-section is to calculate a cryptocurrency demand score from the collected opinion dataset, which will be used to calculate the final reputation of the target entity. Each opinion will be classified into two classes: **high demand** and **low demand** by employing the same architecture of the XLNet classifier model used previously in the paper; except, the classifier will be fine-tuned on a different dataset whereat opinions containing users' purchase intention expressions are collected and then annotated with 'high demand' and 'low demand' labels. Figure 5 shows samples of the training dataset. After classifying each opinion in the dataset based on users' demand for the target entity, a cryptocurrency demand score is calculated using Eq.2, the result is a numerical value between 0 and 0.5 that will be used to compute the final reputation value of the target entity. The max bound value of 0.5 represents the maximum value for the demand score in the final reputation generation formula which will be discussed later in section 4.8.

⁸ <https://www.fool.com/investing/stock-market/market-sectors/financials/cryptocurrency-stocks/value-of-crypto/>

$$DS_j = \frac{HD_j \times 0.5}{HD_j + LD_j} \quad (2)$$

We denote:

DS_j : Demand score toward entity j .

HD_j : Total number of tweets with a high demand toward the target entity j .

LD_j : Total number of tweets with a low demand toward the target entity j .

Review 1 :	[This cryptocurrency has a lot of potential, ^{Expression of purchase intentions} I am buying with no hesitation !]	→	High demand
Review 2 :	[I think I like the use case of this cryptocurrency.]	→	Low demand

Figure 5: Samples from the training dataset used in purchase intention classification

4.5 Popularity score

On social media platforms, customers rely more on the opinions of certain individuals that have some sort of popularity in order to settle on a plan of action toward a specific entity. Researchers argue that trust plays a significant role to determine positive or negative effects on customers' perceptions [LE and Thi 2020]. Also, a research study in ⁹ shows that 72% of customers won't take any buying actions until they've read reviews. In this sub-section, we analyzed each opinion in the dataset for the purpose of distinguishing between tweets with high and low popularity by computing a popularity score for each tweet.

The popularity score is a numerical value between -0.5 and 0.5 calculated based on two selected features of the tweet: likes and shares. The like feature is the number of likes an opinion received from other users, a higher received number of likes implies that the opinion sought approval from other individuals. The second feature is the number of shares an opinion received, a higher received number of shares implies that the opinion is endorsed by other individuals. The popularity score for each tweet is calculated based on the two previously mentioned features (likes and shares) using Eq.3. We multiplied both the number of likes and shares with 0.25 in order to make the psr_{ij} value bounded between 0 and 0.5. In Eq.4, we calculated a custom average of all the opinions' popularity scores in order to obtain a single numerical popularity score PS_j between -0.5 and 0.5 that represents the popularity of the entity j . The PS_j will be used later in the paper to compute the final reputation value. The max bound value of 0.5 represents the maximum value for the popularity score in the final reputation generation formula which will be discussed later in section 4.8.

$$psr_{ij} = \frac{l_{ij} \times 0.25}{\max(L_j)} + \frac{s_{ij} \times 0.25}{\max(S_j)} \quad (3)$$

⁹ <https://mytestimonialengine.com/how-reviews-affect-purchasing-decisions-and-seo>

$$PS_j = \frac{\sum PS_{pos_j} - \sum PS_{neg_j}}{N_j} \quad (4)$$

where:

$$N_j = \begin{cases} Len(PS_{pos_j}), & \text{if } \sum PS_{pos_j} - \sum PS_{neg_j} \geq 0 \\ Len(PS_{neg_j}), & \text{if } \sum PS_{pos_j} - \sum PS_{neg_j} \leq 0 \end{cases}$$

We denote:

- psr_{ij} : Popularity score of tweet i expressed for an entity j .
- l_{ij} : Number of likes received for a tweet i expressed for an entity j .
- s_{ij} : Number of shares received for a tweet i expressed for an entity j .
- $max(L_j)$: Max number of likes received by a tweet toward entity j .
- $max(S_j)$: Max number of likes received by a tweet toward entity j .
- PS_{pos_j} : Set of popularity scores of tweets that describe the entity j positively.
- PS_{neg_j} : Set of popularity scores of tweets that describe the entity j negatively.
- $Len(PS_{pos_j})$: The length of PS_{pos_j} .
- $Len(PS_{neg_j})$: The length of PS_{neg_j} .

4.6 Cryptocurrency news influence score

One of the main aspects of cryptocurrency is that news and events related to a specific cryptocurrency have a significant impact on its reputation. For example, the announcement that Baidu was accepting bitcoins in mid-October 2013 started a surge in its value [Kristoufek and Ladislav 2015]; however, the decision of the Chinese regulation banning the use of bitcoins for electronic purchases in early December 2013 affected the price negatively as well as its reputation in the eye of the buyers. Consequently, we observed that there is a correlation between cryptocurrencies and their related news since whenever breaking good or bad news happens, the price of the cryptocurrencies increases or drops consequently. For this reason, we analyzed the target cryptocurrency's price history in a defined period of time, then we computed a news influence score between -0.5 and 0.5 that will be used for the computation of the reputation value. First, the proposed system scrapes the rate of change (ROC) history of the entity for each day between the publishing date of the oldest and the newest tweet in the collected dataset. ROC is used to mathematically describe the percentage change in value over a defined period of time, and it represents the momentum of a variable, and it is scrapped from a well-known cryptocurrencies price tracker named CoinMarketCap¹⁰. Second, the system determines the highest ROC value of the entire entity's pricing history. Next, the system calculates a news influence value s_{ij} for each day in the defined period of time using Eq.5. Finally, the news influence score NS_j is calculated by averaging all the calculated values s_{ij} using Eq.6.

$$s_{dj} = \frac{roc_{dj} \times 0.5}{max(ROC_j)} \quad (5)$$

¹⁰ <https://coinmarketcap.com/>

$$NS_j = \frac{\sum_{d=otd_j}^{ntd_j} s_{dj}}{n_j} \quad (6)$$

We denote:

roc_{dj} : The percentage of the change in the price of the entity "j" in the day "d" in 24 hours.

$max(ROC_j)$: Max percentage of the change in the price value toward entity j.

otd_j : Publishing date of the oldest tweet toward entity j.

ntd_j : Publishing date of the newest tweet toward entity j.

n_j : Number of days between the publishing date of the oldest and newest tweet toward entity j.

4.7 Reputation generation

The proposed system now has all the necessary elements for the calculation of the final reputation value of an entity. The system aggregates the sentiment score ss_j , the popularity score ps_j , the cryptocurrency demand score ds_j , and finally the cryptocurrency news influence score NS_j in order to produce a single numerical value bounded between 0 and 10 that reflects the reputation of the target entity. Eq.7 displays the calculation of the reputation value for an entity 'j'.

$$Rep_j = max[0, (SS_j + DS_j + PS_j + NS_j)] \quad (7)$$

We denote:

SS_j : sentiment score value for the entity j. This score ranges between 0 and 8.5.

DS_j : demand score value for the entity j. This score ranges between 0 and 0.5.

PS_j : popularity score value for the entity j. This score ranges between -0.5 and 0.5.

NS_j : news influence score value for the entity j. This score ranges between -0.5 and 0.5.

5 Experiments

5.1 Experimental data collection and preprocessing

To evaluate the proposed reputation generation system, we have collected three real-world datasets about three different cryptocurrencies using Twitter API¹¹ and web scraping python scripts. Each dataset contains 250 tweets shared by users expressing their opinions toward the target cryptocurrency. In this paper, we choose to keep the names of the cryptocurrencies being evaluated private, in order to avoid any financial conflict. For this reason, the cryptocurrencies will be referred to as cryptocurrency 1, 2, and 3 for each evaluation dataset 1,2, and 3 respectively. The datasets contain textual tweets with their associated images and emojis. Table 3 shows detailed information about the collected dataset.

We hired three linguistic experts to annotate the sentiment orientation of the tweets (text and image) for the purpose of assessing the performance of our selected sentiment classifier and our proposed sarcasm detection technique. The textual content of each tweet is cleaned and processed by applying the following tasks:

¹¹ <https://developer.twitter.com/en/docs/twitter-api>

	Total number of tweets	Number of positive tweets	Number of negative tweets	Number of neutral tweets	Number of sarcastic tweets
Dataset 1	250	168	38	44	31
Dataset 2	250	92	111	47	20
Dataset 3	250	101	89	60	53

Table 3: Information about the evaluation datasets

- Removing special characters and duplicated words.
- Eliminating attached links.
- Replacing slang words.
- Preparing the textual opinion for the XLNet classifier (tokenization, special tokens addition, etc).

We have exploited an up-to-date list of slang words and abbreviations used in today's modern tweets and chat texts¹². The list was extracted from multiple sources (The Urban Dictionary, The Most Used Internet Abbreviations for Texting and Tweeting, SMS Slang Translator, English abbreviations). The list was built in the form of a dictionary that contains slang words with their full form.

5.2 Sentiment classification evaluation

The XLNet model was employed for the sentiment classification task, where it was fine-tuned on the SST dataset [Socher et al. 2013]. Since the sentiment orientation of the tweets could be positive, negative, or neutral, we have used SST-5 dataset which is labeled as either negative, somewhat negative, neutral, somewhat positive, or positive. We automatically labeled negative and somewhat negative reviews as negative (label 0), neutral reviews as neutral (label 1), and positive and somewhat positive reviews as positive (label 2). We refer to the processed dataset as SST-3 since it contains 3 labels. The training set consists of 3738 negative reviews, 1853 neutral reviews, and 4054 positive reviews, as for the test set, it contains 912 negative reviews, 389 neutral reviews, and 909 positive reviews. The classifier was built using the Hugging-Face library¹³, and its parameters are described in Table 4.

Hyperparameters	Settings
Epochs	4
Learning rate	1e-5
Batch size	8
Optimizer	Adam [Kingma and Ba. 2015]

Table 4: Fine-tuned XLNet model's parameters

¹² <https://www.kaggle.com/code/nmaguette/up-to-date-list-of-slangs-for-text-preprocessing/notebook>

¹³ <https://huggingface.co/>

We have fine-tuned XLNet on the SST-3 dataset because (1) SST-2 dataset is basically SST-5 dataset without neutral reviews (negative or somewhat negative vs somewhat positive or positive with neutral sentences discarded), and (2) XLNet achieved excellent results on the SST-2 dataset as shown in Table 5. For the SST-3 test set, XLNet achieved 82.40 %, 28.72%, and 86.80 % f1-score respectively for detecting negative, neutral, and positive reviews.

Classifier	Accuracy (%)
CNN	87.2
Bi-LSTM	87.5
BERT-base	91.2
XLNet	97.0

Table 5: Accuracy comparison between XLNet model and previous classifier on the SST-2 dataset according to Papers With Code¹⁴ leaderboard

5.3 Sarcasm detection evaluation

After extracting the sentiment orientation of each textual content of the tweets in the three evaluation datasets, the system performs the sentiment classification for the images associated with the tweets using the fine-tuned ResNet50 classifier. Tweets that do not contain any images will not go through this process. The sentiment orientation of the emojis associated with the textual content is also extracted. After acquiring the sentiment of the texts, the sentiment of the images, and the sentiment of the emojis, the proposed system applies the rules previously mentioned in Table 2 to identify sarcastic tweets. Figure 6 and Table 8 present the accuracy of the proposed sarcasm detection technique with and without using the sarcasm detection phase. As we can see, by employing the sarcasm detection technique, the sentiment orientation task achieve good performance. This means that the proposed sarcasm detection technique helps uncover the true sentiment of the tweets.

In order to investigate the impact of the sarcasm detection phase on text-only tweets, we have conducted two experiments. In the first experiment, we fine-tuned the XLNet model on the SST-3 training set. After that, we fine-tuned BERTweet-base [DQ Nguyen et al. 2020] model on iSarcasmEval training set that was proposed in SemEval-2022 Task 6 [IA Farha et al. 2022] (the model was ranked 8th in the competition [Benlahbib et al. 2022]), then we applied it on the SST-3 test set. We set two rules, (1) if the review is classified as positive by XLNet and classified as sarcastic by BERTweet-base, then it will be labeled as negative, and (2) if the review is classified as negative by XLNet and classified as sarcastic by BERTweet-base, then it will be labeled as positive. Table 6 depicts the result of XLNet on the SST-3 test set before and after applying the sarcasm detection phase.

In the second experiment, we fine-tuned the BERTweet-base model on the Tweet Sentiment Extraction¹⁵ training set. After that, we fine-tuned the same model on iSarcasmEval training set, then we applied it to the Tweet Sentiment Extraction test set. We

¹⁴ <https://paperswithcode.com/sota/sentiment-analysis-on-sst-2-binary>

¹⁵ <https://www.kaggle.com/competitions/tweet-sentiment-extraction/data?select=test.csv>

Model	Precision	Recall	F1-score	Accuracy
XLNet without sarcasm detection	0.6660	0.6775	0.6597	0.7715
XLNet with sarcasm detection	0.6630	0.6749	0.6571	0.7679

Table 6: Comparison between the sentiment orientation results with and without employing the sarcasm detection technique on the SST-3 test set

set two rules, (1) if the tweet is classified as positive and sarcastic, then it will be labeled as negative, and (2) if the tweet is classified as negative and sarcastic, then it will be labeled as positive. Table 7 depicts the result of BERTweet-base on the Tweet Sentiment Extraction test set before and after applying the sarcasm detection phase.

Model	Precision	Recall	F1-score	Accuracy
BERTweet-base without sarcasm detection	0.8114	0.8047	0.8072	0.8039
BERTweet-base with sarcasm detection	0.7713	0.7739	0.7671	0.7663

Table 7: Comparison between the sentiment orientation results with and without employing the sarcasm detection technique on the Tweet Sentiment Extraction test set

We notice from Tables 6 and 7 that the sentiment classification performance decreased after applying the sarcasm detection phase. Therefore, we have decided not to use a sarcasm detection phase for text-only tweets since it is hard to detect implicit sarcasm from textual content even for humans due to the need for pragmatic analysis.

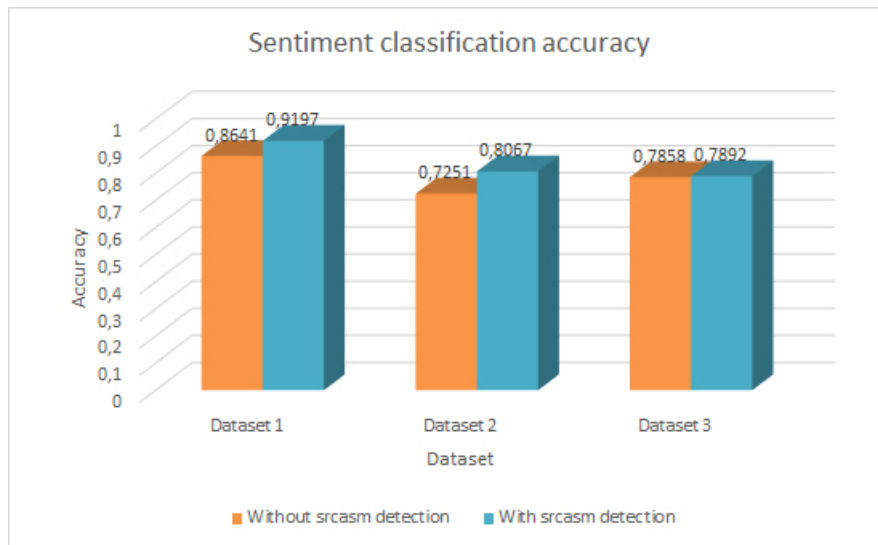


Figure 6: Accuracy of the sentiment classification on the evaluation datasets with and without the proposed sarcasm detection technique

		True sentiment orientation (%)	Sentiment orientation results (%) without the sarcasm technique	Sentiment orientation results (%) with the sarcasm technique
Dataset 1	Positive	67.2	71.1	69
	Negative	15.2	10.5	12.6
	Neutral	17.6	18.4	18.4
Dataset 2	Positive	36.8	50.77	40.1
	Negative	44.4	38.12	48.77
	Neutral	18.8	11.13	11.13
Dataset 3	Positive	40.4	51.09	46.5
	Negative	35.6	34.69	39.28
	Neutral	24	14.22	14.22

Table 8: Comparison between the sentiment orientation results after employing the sarcasm detection technique on the three evaluation datasets

5.4 Reputation system evaluation

The proposed reputation generation system is the first in the literature to produce the reputation of cryptocurrencies based on peoples' tweets shared on Twitter. The system employs various technologies and features such as the sentiment orientation of the tweets, and their popularity, as well as taking into consideration the aspect of demand for a specific cryptocurrency, and the influence of news and events. Therefore, in order to evaluate the effectiveness of the proposed system, and due to the nonexistence of standard evaluation metrics for this kind of task. We sought help from a well-known cryptocurrency website named Cryptobubbles¹⁶ by asking its Twitter followers to give their thoughts in form of numerical ratings toward three cryptocurrencies related to the three evaluation datasets described in Table 3. Thus, we managed to collect 827 ratings for each cryptocurrency that reflect its reputation in the eye of the users based on their knowledge and experience, and also on the overall features of the target cryptocurrency: related news, public sentiment, etc. The goal of this experiment is to compare the average of all ratings given by the users, which is considered the ground truth, with the output value produced by the proposed reputation system. In table 9, we calculated the average of all the given ratings for each cryptocurrency related to our evaluation datasets using $\mu = \frac{1}{n} \sum_{i=0}^n x_i$, where $x_1, x_2, \dots, x_n$ are the collected ratings. We also calculated the standard deviation, which is a measure of the amount of variation or dispersion of a set of values. The standard deviation formula is $\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2}$ where $x_1, x_2, \dots, x_n$ are the observed values (ratings) of the sample items, μ is the mean value of these observations, and n is the number of observations in the sample. We notice that the standard deviation value for each cryptocurrency is very low, which means that the ratings given for each cryptocurrency are balanced.

In Figure 9 we compared the users' average ratings (ground truth) and the proposed system output reputation value. we notice that our system was able to produce a very close numerical reputation value to the ground truth. We calculated the absolute error of the system's outputs $AE = |V_a - V_e|$, where V_a is the measured value (the proposed system's output) and V_e is the exact value (users' average ratings). The AE for Dataset 1 is '0.5', for dataset 2 the AE is '0.74', and for dataset 3 the AE is '0.17'. We notice that AE is very low for all three datasets which means that the proposed system is consistent when generating the reputation values. Based on the experiment results, we observed

¹⁶ <https://cryptobubbles.net/>

	Users' average rating	Standard deviation
Cryptocurrency 1	7.5	0.53
Cryptocurrency 2	4.0	0.67
Cryptocurrency 3	8.5	0.46

Table 9: Users' average ratings and standard deviation toward the three cryptocurrencies

that the proposed system was able to produce an accurate reputation value from only 250 tweets compared with the average rating of 827 users.

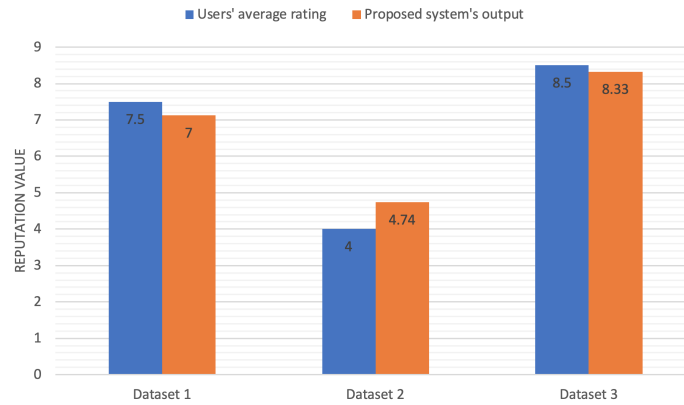


Figure 7: Comparison between the system's output reputation value and users' average rating (Ground truth)

5.5 Ablation study

In this sub-section, we analyze the impact of the features employed in the proposed reputation system. As mentioned previously, four features (Sentiment score, cryptocurrency demand feature, popularity feature, and news influence feature) were used as a basis for the purpose of generating an accurate reputation value. We conducted an ablation study and demonstrated the results in Table 10. We can observe from Table 10 that the system which exploited neither "popularity feature" nor "news influence feature" achieve poor results on all three datasets. Comparatively, the system which doesn't consider "popularity feature" and "cryptocurrency demand feature" performs better. This shows that the "news influence feature" provides useful insight to the system for more accurate reputation computing. In addition, disregarding the "cryptocurrency demand feature" (System w/o d), and "news influence feature" (System w/o n) leads to poor performance. We can conclude that both, the "news influence feature" and "demand feature" are important to maximize the performance of the proposed reputation system. We can also notice that neglecting the "popularity feature" (System w/o p) reduces the performance of the model slightly, which broadly indicates the importance of identifying

Models	Dataset 1	Dataset 2	Dataset 3
	Percent error (%)	Percent error(%)	Percent error(%)
sytem w/o p+n	9.33	25	5.88
sytem w/o p+d	7.13	21.2	3.79
sytem w/o n	7.88	22.4	4.6
system w/o d	6.91	18.78	2.6
system w/o p	6.98	18.82	2.44
Cryptocurrency reputation system	6.66	18.5	2.0

Table 10: Experimental results of ablation study on the three evaluation datasets. “p” represents popularity feature, “d” represents cryptocurrency demand feature, “n” represents news influence feature

the popularity of the tweets toward the target cryptocurrency. The formula of the percent error is computed in the following way: $Percent_Error = \frac{measured - real}{real} \times 100\%$

6 Conclusions and Future Work

In this paper, we proposed a reputation system capable of generating numerical reputation values for a specific cryptocurrency based on tweets. The contribution of this work revolves around three components that were not exploited in previous systems. The first one is that our proposed system is considered the first one to handle the reputation generation of cryptocurrencies. The second one is sarcasm detection, where sarcastic opinions are detected and treated for the purpose of extracting the authentic sentiment of those opinions. The third one is considering the intensity of demand for that specific entity as well as taking into consideration the influence of news upon the target entity for accurate reputation generation. Then, we incorporated the previous results with a calculated popularity score using mathematical formulas to finally obtain a reputation value for the targeted entity. We conducted an experiment on three real-world Twitter datasets related to three cryptocurrencies, where we compared the proposed system's output results with the ground truth, which is the average votes of 827 Twitter users expressing their satisfaction toward three cryptocurrencies by giving a numerical score between 0 and 10. The comparison shows that our system produces a reliable reputation value that could be used in real-life applications. Our system can serve as a decision-making tool for users and experts in order to measure the reputation of cryptocurrencies. Therefore, our future studies will focus on generating more than just the numerical reputation value, instead, we will try to extend the system to generate a textual summary of the pros and cons of the targeted cryptocurrency. Also, we intend to extend this system to be used for different types of financial assets.

References

- [Abdel-Hafez et al. 2012] A. Abdel-Hafez, Y. Xu, and D. Tjondronegoro, “Product reputation model: An opinion mining based approach,” in Proceedings of the 1st International Workshop on Sentiment Discovery from Affective Data [CEUR Workshop Proceedings Volume 917], pp. 16–27, 2012.
- [Abu-Bakar 2018] Abu-Bakar Mufti Muhammad, “Shariah analysis of bitcoin, cryptocurrency, and blockchain”. Shariah Analysis in Light of Fatwas and Scholars’ Opinions. 2018.

- [Behdenna et al. 2018] Behdenna Salima, Fatiha Barigou, and Ghalem Belalem, "Document level sentiment analysis: a survey." *EAI Endorsed Transactions on Context-aware Systems and Applications* 4.13 (2018).
- [Benlahbib and Nfaoui 2019] A. Benlahbib and E. H. Nfaoui, "An Unsupervised Approach for Reputation Generation," *Procedia Computer Science*, vol. 148, pp. 80-86, 2019.
- [Benlahbib and Nfaoui 2019] A. Benlahbib and E. H. Nfaoui, "A hybrid approach for generating reputation based on opinions fusion and sentiment analysis," *J. Organ. Comput. Electron. Commer.*, pp. 1–19, 2019.
- [Benlahbib et al. 2019] A. Benlahbib, A. Boumhidi, and E. H. Nfaoui, "A Logistic Regression Approach for Generating Movies Reputation Based on Mining User Reviews," in *International Conference on Intelligent Systems and Advanced Computing Sciences (ISACS)*, Taza, Morocco, pp. 1–7, 2019.
- [Benlahbib and Nfaoui 2020] A. Benlahbib and E. H. Nfaoui, "Aggregating Customer Review Attributes for Online Reputation Generation," *IEEE Access*, vol. 8, pp. 96550–96564, 2020.
- [Benlahbib and Nfaoui 2020] A. Benlahbib and E. H. Nfaoui, "AmazonRep: A Reputation System to Support Amazon's Customers Purchase Decision Making Process based on Mining Product Reviews," *2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS)*, 2020, pp. 1-8, doi: 10.1109/ICDS50568.2020.9268769.
- [Benlahbib and Nfaoui 2021] A. Benlahbib and E. H. Nfaoui, "MTVRep: A movie and TV show reputation system based on fine-grained sentiment and semantic analysis," *Int. J. Electr. Comput. Eng. IJECE*, vol. 11, Art. no. 2, 2021.
- [Benlahbib et al. 2022] A. Benlahbib, H. Alami, and A. Alami, "LISACTeam at SemEval-2022 Task 6: A Transformer based Approach for Intended Sarcasm Detection in English Tweets." In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 993–998, Seattle, United States. Association for Computational Linguistics.
- [Benlahbib et al. 2022] A. Benlahbib, A. Boumhidi, and E. H. Nfaoui, "Mining Online Reviews to Support Customers' Decision-Making Process in E-Commerce Platforms: a Narrative Literature Review." *Journal of Organizational Computing and Electronic Commerce*, 2022, doi: 10.1080/10919392.2022.2053454.
- [Boumhidi and Nfaoui 2020] A. Boumhidi and E. H. Nfaoui, "Mining User's Opinions and Emojis For Reputation Generation Using Deep Learning." *2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS)*. IEEE, 2020.
- [Boumhidi et al. 2021] A. Boumhidi, A. Benlahbib, and E. H. Nfaoui, "Mining Online Opinions and Reviews Using Bi-LSTM for Reputation Generation," in *Artificial Intelligence and Industrial Applications*, pp. 135–144, 2021.
- [Boumhidi and Nfaoui 2021] A. Boumhidi and E. Nfaoui, "Leveraging lexicon-based and sentiment analysis techniques for online reputation generation," *International Journal of Intelligent Engineering and Systems*, 2021.
- [Boumhidi et al. 2022] A. Boumhidi, A. Benlahbib, and E. H. Nfaoui, "Cross-Platform Reputation Generation System based on Aspect-based Sentiment Analysis." *IEEE Access* (2022).
- [Clark et al. 2020] Clark, K., Luong, M. T., Le, Q. V., & Manning, C. D., "Electra: Pre-training text encoders as discriminators rather than generators." *arXiv preprint arXiv:2003.10555* (2020).
- [Croman et al. 2016] Croman, Kyle and Decker, Christian and Eyal, Ittay and Gencer, Adem Efe and Juels, Ari and Kosba, Ahmed and Miller, Andrew and Saxena, Prateek and Shi, Elaine and Gün Sirer, Emin and Song, Dawn and Wattenhofer, Roger", editor="Clark, Jeremy and Meiklejohn, Sarah and Ryan, Peter Y.A. and Wallach, Dan and Brenner, Michael and Rohloff, Kurt, "On scaling decentralized blockchains." *International conference on financial cryptography and data security*. Springer, Berlin, Heidelberg, 2016.

- [Dai et al. 2019] Dai, Z., Yang, Z., Yang, Y., Carbonell, J., Le, Q. V., & Salakhutdinov, R., "Transformer-xl: Attentive language models beyond a fixed-length context." arXiv preprint arXiv:1901.02860 (2019).
- [Devlin et al. 2018] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K., "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).
- [DQ Nguyen et al. 2020] Dat Quoc Nguyen, Thanh Vu, and Anh Tuan Nguyen, "BERTweet: A pre-trained language model for English Tweets." In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pages 9–14, Online. Association for Computational Linguistics. 2020.
- [Elmurngi and Gherbi 2020] E. I. Elmurngi and A. Gherbi, "Building sentiment analysis model and compute reputation scores in e-commerce environment using machine learning techniques," International Journal of Organizational and Collective Intelligence (IJOIC), vol. 10, no. 1, pp. 32–62, 2020.
- [Gupta et al. 2020] A. Gupta, S. Priyani, and R. Balakrishnan, "Customized reputation generation of entities using sentiment analysis," World Journal of Engineering, 2020.
- [He et al. 2016] He, K., Zhang, X., Ren, S., & Sun, J., "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- [IA Farha et al. 2022] Ibrahim Abu Farha, Silviu Vlad Oprea, Steven Wilson, and Walid Magdy, "SemEval-2022 Task 6: iSarcasmEval, Intended Sarcasm Detection in English and Arabic." In Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022), pages 802–814, Seattle, United States. Association for Computational Linguistics. 2022.
- [Kingma and Ba. 2015] Kingma, Diederik P. and Ba, Jimmy, "Adam: A Method for Stochastic Optimization." In 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings.
- [Kristoufek and Ladislav 2015] Kristoufek, Ladislav, "What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis." PloS one 10.4 (2015): e0123923.
- [Lan et al. 2019] Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R., "Albert: A lite bert for self-supervised learning of language representations." arXiv preprint arXiv:1909.11942 (2019).
- [LE and Thi 2020] LE, Nguyen Binh Minh, and Thi Phuong Thao HOANG, "Measuring trusts and the effects on the consumers' buying behavior." Journal of Distribution Science 18.3 (2020): 5-14.
- [Liu et al. 2015] Liu, X., Gao, J., He, X., Deng, L., Duh, K., & Wang, Y. Y., "Representation learning using multi-task deep neural networks for semantic classification and information retrieval." (2015).
- [Liu et al. 2019] Liu, X., He, P., Chen, W., & Gao, J., "Multi-task deep neural networks for natural language understanding." arXiv preprint arXiv:1901.11504 (2019).
- [Liu et al. 2019] Yinhan Liu and Myle Ott and Naman Goyal and Jingfei Du and Mandar Joshi and Danqi Chen and Omer Levy and Mike Lewis and Luke Zettlemoyer and Veselin Stoyanov, "A robustly optimized bert pretraining approach." arXiv preprint arXiv:1907.11692 (2019).
- [Raffel et al. 2019] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Li, W. & Liu, P. J., "Exploring the limits of transfer learning with a unified text-to-text transformer." arXiv preprint arXiv:1910.10683 (2019).
- [Rajpurkar et al. 2016] Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P., "Squad: 100,000+ questions for machine comprehension of text." arXiv preprint arXiv:1606.05250 (2016).
- [Resnick and Zeckhauser 2002] P. Resnick and R. Zeckhauser, "Trust among strangers in internet transactions: Empirical analysis of eBay's reputation system," in Advances in Applied Microeconomics, vol. 11, pp. 127–157, 2002.

[Schneider et al. 2000] J. Schneider, G. Kortuem, J. Jager, S. Fickas, and Z. Segall, "Disseminating trust information in wearable communities," *Pers. Technol.*, vol. 4, no. 4, pp. 245–248, 2000.

[Sejung and Han Woo. 2021] Park, Sejung, and Han Woo Park, "Relations between Reputation and Social Media Marketing Communication in Cryptocurrency Markets: Visual Analytics Using Tableau." *International Journal of Contents*, vol. 17, no. 1, Mar. 2021, pp. 1–10, doi:10.5392/IJOC.2021.17.1.001.

[Socher et al. 2013] Socher, Richard and Perelygin, Alex and Wu, Jeon and Chuang, Jason and Manning, Christopher D. and Ng, Andrew and Potts, Christopher, "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank." *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 2013.

[Vaswani et al. 2017] Ashish Vaswani and Noam Shazeer and Niki Parmar and Jakob Uszkoreit and Llion Jones and Aidan N. Gomez and Lukasz Kaiser and Illia Polosukhin, "Attention is all you need." *Advances in neural information processing systems* 30 (2017).

[Wang et al. 2018] Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R., "GLUE: A multi-task benchmark and analysis platform for natural language understanding." *arXiv preprint arXiv:1804.07461* (2018).

[Yan et al. 2017] Z. Yan, X. Jing, and W. Pedrycz, "Fusing and mining opinions for reputation generation," *Inf. Fusion*, vol. 36, pp. 172–184, 2017.

[Yang et al. 2019] Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V., "Xlnet: Generalized autoregressive pretraining for language understanding." *Advances in neural information processing systems* 32 (2019).