International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys.13, No.1 (Mar-23)

http://dx.doi.org/10.12785/ijcds/130137

The Influence of Social Media on Cryptocurrency Price: A Sentiment Analysis Approach

Bousselham EL HADDAOUI[1], **Raddouane CHIHEB**[1], **Rdouan FAIZI**[1] **and Abdellatif EL AFIA**[1]

[1]*ENSIAS, MOHAMMED V UNIVERSITY IN RABAT, MOROCCO*

Received 15 Feb. 2022, Revised 13 Jan. 2023, Accepted 5 Mar. 2023, Published 16 Mar. 2023

Abstract: Decentralized Web (Web3) and Finance (DeFi) have become the main discussion topic in research and industry fields. Cryptocurrencies, as an essential part of DeFi, enjoyed the interest of many stakeholders such as companies, professionals, researchers, and even common citizens eager to benefit from the proposed ecosystems. Although previous research studies focused on establishing price prediction systems using Sentiment Analysis (SA) techniques, the main focus of these studies was the performance of the predictive system rather than the accuracy and efficiency of the used models. In our work, we address two research questions; the predictability of cryptocurrency price based on past social and technical information, and the effect of social features on cryptocurrency price fluctuations using an SA and a Time Series approach. A combination of selected social and technical features was processed and reframed as a prediction problem, then studied to assess the ability of our model to predict the desired price. We noted that there is both an explicit correlation for some considered features and implicit for others, also social features including overall positive and neutral sentiment, and community engagement improved the performance of our model.

Keywords: Cryptocurrencies, Social Media, Sentiment Analysis, Time Series Prediction, Transformers, LSTM

1. INTRODUCTION

The cryptocurrency ecosystem is surging with a closing market cap of \$2 T[1] in 2021, a recent market performing among the top such as the gold market at \$11.57 T[2] . Bitcoin [1] blockchain is one of the dominant cryptocurrencies with 45.3% shares of the crypto market, its innovative technology is driving governments and industry's recent adoption as a form of new money. Moreover, new blockchain concepts are emerging, seeking more scalability, security, and decentralization, thus providing a robust foundation for various use cases.

Cryptocurrencies have become an important component of the global financial system alongside traditional stocks. These ecosystems have notable disparities that shifted the attention from a stock market characterized by ownership of real assets, reliability, regulated, and a difficult entry-level to a new form of investment, which is the cryptocurrency market, bringing innovation and freedom constrained with an unregulated market, high-risk investment, backed with popularity and technological aspects, and accessible for everyone.

The volatility [2] of cryptocurrencies' prices, especially bitcoin, is driving the need to elaborate techniques enabling the analysis and the prediction of its behavior with the same maturity as in stocks. In this respect, current approaches focus on extracting features that have predictive power from market, technical and social information. The market and technical features are closely related to cryptocurrency price; thus, they hold information that directly affects the price variations. However, social information including the volume, the overall sentiment, influencers' interactions, and community engagement is difficult to process. Sentiment Analysis (SA) is "the field of study that analyzes people's opinions, sentiments, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes" as defined in [3], helps in the process of extracting information from social information to study its effect on cryptocurrency price.

Social media influence on cryptocurrencies price fluctuations has been vaguely investigated due to the difficulty in representing certain social aspects such as community interactions, content engagement, etc. and which features have a direct and strong impact on the cryptocurrency price. Moreover, social features should be linked with technical features that are available per fixed time intervals, thus limiting the training dataset for the price prediction models. Current approaches usually use similar systems such as SA and time series prediction models. However, they focus only

[1]https://www.coingecko.com
[2]https://8marketcap.com

E-mail address: bousselham.haddaoui@um5s.net.ma, r.chiheb@um5s.net.ma, r.faizi@um5s.net.ma, a.elafia@um5s.net.ma

on a sub-system to improve or validate hypotheses, and the other systems are often ready to use state-of-the-art or baseline systems for evaluation purposes.

In this paper, we propose a cryptocurrency price prediction system to answer two research questions. First, is the ability of past social and technical information to predict cryptocurrency price fluctuations. Second, identifying the social features which have a strong influence on cryptocurrencies price prediction. The rest of the paper is presented as follows: in section 2 we highlight related work and important research studies. Section 3 will describe the proposed system, its architecture, and various components along with the dataset. Section 4 will present the key findings of our study and the discussion of the effectiveness of the proposed system. Finally, section 5 will present conclusions and future work.

2. RELATED WORK

Since the emergence of bitcoin in 2009, various studies have investigated the predictability of its price although the cryptocurrency market is known to be extremely volatile. The efficient market hypothesis EMH [4] and the Random Walk, the foundations of stock market price prediction, assume that the price is influenced only by future information. However, the following studies questioning these foundations affirmed that early indicators for price prediction can be extracted from past information [5]. Whereas market and technical information items are the first indicators directly related to cryptocurrency price variations, additional indicators can be emotions and mood states which are important for financial decision making according to behavioral finance [6]. These social attributes are available through sentiment analysis techniques, which usually retrieve content's specific or overall sentiment into categories such as: positive, negative, or neutral. Furthermore, additional social attributes can be considered essentially content volume and community interactions [7].

The correlation between market, technical and social features and the crypto price has been investigated in previous studies [8] [9], public mood, google search and exchange volume variations hold predictive information about cryptocurrency price. Moreover, the correlation of market sentiment features including investor's sentiment, supply and demand, cost of mining, government regulations, and external events such as economic crisis with cryptocurrency price was explored using Pearson's correlation coefficient [10]. Further studies discussed the influence of social content sentiment on price variations, yet the degree of influence is still investigated. In [7], the authors stated that cryptocurrency sentiment is less effective when the price is falling as most social content is objective or holds a positive orientation regardless of the price change direction. However, neutral tweets were proved to be the larger part of social content and convey information that influences the price [2]. While a relationship can be established between cryptocurrency price and sentiment, the latter has less impact when there's an abnormal price rise or fall movement. Another perspective to consider is the size of transactions as it was observed in [11] that small transactions allow better price prediction rather than large transactions. Moreover, large transactions are usually related to events leading to sudden and notable price changes and usually propagate to generate a small transaction trend.

Social features processing is performed efficiently using Sentiment Analysis techniques and benefits from an orchestration of natural processing language (NLP) tasks such as data acquisition, preprocessing, classification, and visualization [12]. Current research trends for applied SA to the cryptocurrency domain rely on ready-to-use classification tools or baseline models that deliver average performance; thus, the impact of social features is less accurate. In the data acquisition and preprocessing phrases, most of the studies rely on Twitter, Reddit, news, or google research trends information as data sources. Since the crypto community uses mostly similar language structures across different social channels, the proposed models aren't sensitive to data sources [13]. In addition, social content lacks structure and contains high levels of noise which bring more complexity to SA models. Various preprocessing methods are used, as needed, for content preprocessing that preserves sentence meaning and removes unnecessary such as tokenization, lemmatization, normalization (remove URLs, white spaces, user mentions, remove "RT" in retweets, remove tweets with less than 4 tokens, the hashtag '#' is removed if the word is present in an English dictionary, otherwise the whole hashtag is removed, expand word contractions (i.e. "we're" becomes "we are"), lowercase text, handle negation, remove ticker symbols (i.e. $BTC), remove slang, repeated letters, punctuation, and stop words) [14] [10].

In the classification phase, earlier studies investigated Machine Learning (ML) based models such as Naive Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), K-Means, Linear Regression (LR), Decision Trees (DT) [15] [16] [17]. Moreover, genetic algorithms have been used for algorithm selection along with ensemble techniques [11]. Following this, a research trend was observed in using available and ready-to-use tools such as VADER [18], OpinionFinder [19], and Google-Profile for Mood States (GPOMS) due to their reasonable performance and straightforward integration. However, possible gains from social attributes were limited after the emergence of Deep Learning (DL) techniques which provided considerable performance enhancement compared to machine learning algorithms [20]. Neural networks such as Recurrent Neural Networks (RNN), especially the Long Short-Term Memory (LSTM) variant, and Convolutional Neural Networks (CNN) combined with word embedding techniques such as Global Vectors (GloVe), FastText, or Word2Vec achieved better results in the classification of social content [21] [22]. Current DL models provide the ability to model language complexity and capture high-level patterns to overcome machine learning algorithms' limitations. However,

Date	Open	High	Low	Close	Volume	Transactions	Hash rate
2021-01-01	28994.009766	29600.626953	28803.585938	29374.152344	40730301359	258080	137764027.0294169
2021-01-02	29376.455078	33155.117188	29091.181641	32127.267578	67865420765	297111	139613208.60028157
2021-01-03	32129.408203	34608.558594	32052.316406	32782.023438	78665235202	359116	146085344.09830785
2021-01-04	32810.949219	33440.21875	28722.755859	31971.914063	81163475344	373734	159954205.87979275
2021-01-05	31977.041016	34437.589844	30221.1875	33992.429688	67547324782	354091	145160753.31287554
2021-01-06	34013.613281	36879.699219	33514.035156	36824.363281	75289433811	397384	163652569.02152207
2021-01-07	36833.875	40180.367188	36491.191406	39371.042969	84762141031	401744	155331251.95263115
2021-01-08	39381.765625	41946.738281	36838.636719	40797.609375	88107519480	358526	136839436.24398455
2021-01-09	40788.640625	41436.351563	38980.875	40254.546875	61984162837	321389	133808024.52210464
2021-01-10	40254.21875	41420.191406	35984.628906	38356.441406	79980747690	331865	162879650.9247733

Table I. TECHNICAL BITCOIN DATA FROM 1/1/2021 TO 18/6/2021

Date	Total Tweets	Replies Count	Retweets Count	Likes Count	Positive	Negative	Neutral
2021-01-01	8107	7180	12853	45180	503	154	7449
2021-01-02	17666	11743	15042	100367	1117	385	16164
2021-01-03	16165	9071	14089	72697	1073	346	14745
2021-01-04	6258	4905	10709	30771	433	178	5647
2021-01-05	12323	8119	10450	50157	845	315	11163
2021-01-06	15877	8962	12407	66146	1045	306	14525
2021-01-07	20941	24710	27063	106865	1371	428	19142
2021-01-08	19195	39303	42979	104353	1074	418	17703
2021-01-09	13920	8809	11590	53790	698	240	12982
2021-01-10	13458	8867	13980	60952	720	377	12361

Table II. SOCIAL BITCOIN DATA FROM 1/1/2021 TO 18/6/2021

they require considerable training time and large labeled training data which isn't usually available, especially in the context of financial social content as the community tends to use a specific domain language. In this respect, Transformers [23] were investigated for use in financial sentiment analysis systems to leverage language complexity based on pre-trained models fine-tuned using transfer techniques. Moreover, various pre-trained transformer models were considered such as: BERT, XLNet, XLM, ALBERT, RoBERTa, XLM-RoBERTa, and BART [24]. Domain fine-tuned models were also made available such as FinBERT [25] to achieve both gains in performance and ready-to-use models that are time and cost-efficient.

A similar research path was conducted to address the price prediction problem which was considered a classical time series forecasting problem. In fact, deep learning models, especially LSTM networks [26] [27], were considered due to the non-linearity of the problem and their effectiveness compared to machine learning models such as SVM [28]. The ability to predict the cryptocurrencies price relies on features correlation, stationarity, and seasonality aspects [22], thus specific tests are carried out to reframe the selected features to suit a time series problem. In [29], a comprehensive survey was conducted where the authors stated stationarity graphical (i.e., correlogram, covariogram, etc.) and statistical (i.e., short-time Fourier Transform, Wavelet Transform) tests alongside time-domain methods for seasonality tests (i.e., unit root, breakpoint analysis, etc.). In addition, the authors provided various methods for time series prediction problems such as linear (i.e. Linear Autoregressive (AR), Moving Average (MA) and Autoregressive Moving Average (ARMA)), non-linear (i.e. Polynomial Autoregressive Model (PAR), Functional-coefficient Autoregressive Model (FAR), Markov Switching Autoregressive Models (MSAR), Smooth Transition Autoregressive Models (STAR) and Autoregressive Conditional Heteroscedasticity (ARCH)) and deep learning (i.e. Recurrent Neural Networks (RNN), Long Short-Term Memory Networks (LSTM), Convolutional Neural Networks (CNN) and Transformers) time series models. Moreover, time series presents many challenges, essentially the window size and prediction timesteps. The window size and the prediction timestamps depend heavily on the training data and the model's quality, in [2] the window size was fixed in days to study the optimal prediction timesteps. However, both the parameters can be fixed to appreciate the model quality [26].

3. METHODOLOGY

The proposed methodology considers lagged values of different bitcoin-related features to forecast the next day's price. In this respect, we gather BTC blockchain information and related tweets and proceed to necessary preprocessing tasks before feeding the dataset to the model for training and evaluation.

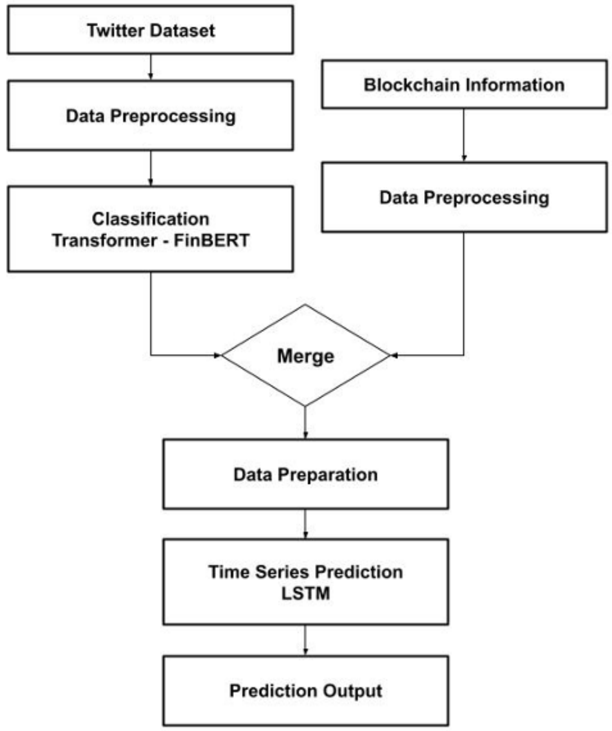

Figure 1. Bitcoin price prediction system architecture

A. Dataset and Preprocessing

Since cryptocurrency is an online investment, most projects rely on social media for their communication campaigns. Moreover, social media users or groups of special interest spread information or organize campaigns that may influence investors buying or selling decisions and cause a price up or downtrend. Besides, blockchain information such as the daily volume, the number of transactions, the difficulty of mining, etc. may hold direct information that influences price fluctuations. In this respect, we gathered various social and technical features that may directly influence the price value. We scraped 3 214 241 bitcoin-related tweets using Twint[3] , an open-source python tool that help retrieve tweets using custom queries (i.e., #btc in our case). We also gathered bitcoin blockchain information from blockchain.com, which is a reliable source of getting specific technical information such as open, high, low, and close (OHLC) BTC price, volume, number of transactions, and the hash rate. Social and technical data were retrieved for the period from 1 January 2021 to 18 June 2021.

Social data are very noisy with crypto-related language, mentions, random emoticons, numbers, and punctuations, etc. The non-conventional used structure is an additional complexity to consider when dealing with social media content, thus a preprocessing step before analysis is recommended. We performed a content cleaning using a set of commonly [14] used preprocessing tasks in social media context such as: removing URLs, mentions, numbers, punctuation, special characters, and formatting the content to a lowercase text structure. Moreover, we aggregated social engagement feature values such as tweets volume, replies, replies and likes count together with social overall sentiment state encoded in the positive, negative, and neutral tweets count that are results of the classification process that will be explained in the section 3-B1.

Blockchain technical data are available and require less preprocessing effort compared to social data. Each available technical factor data was gathered and then aggregated by day. Table I shows an overview of technical bitcoin data, and Table II shows an overview of processed social bitcoin data for the selected period.

B. System Architecture

The proposed system in Figure 1 performs two main steps: tweet classification and bitcoin price forecasting. A classification process is carried out following the preprocessing of the collected social data, and the collected social and technical data are reframed to address a time series forecasting problem.

1) The classification model

We used a transformer-based approach to classify social data due to the fascinating results and the notable ability to generalize on NLP-related tasks. Since the introduction of the Bidirectional Encoder Representations from Transformers - BERT [23], language modeling techniques

[3]https://github.com/twintproject/twint

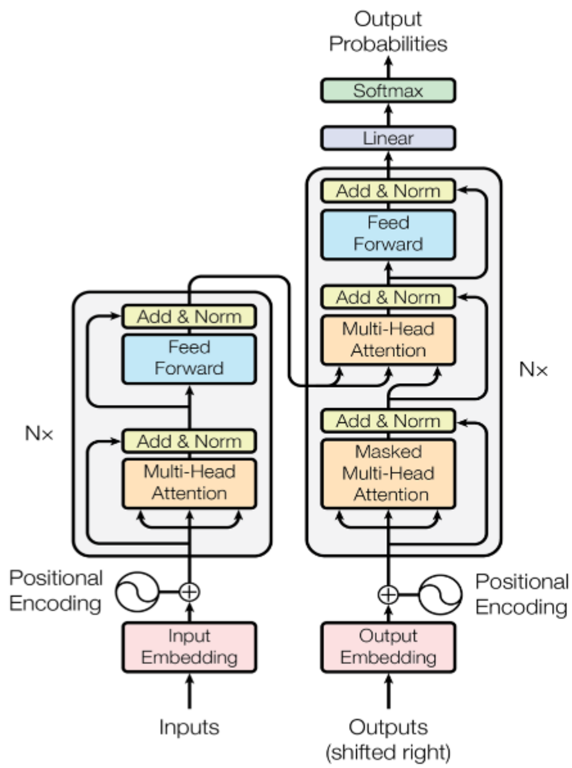

Figure 2. The Transformer model architecture [23]

redirected the focus from model architecture complexity to pre-training and fine-tuning to specific domains. BERT model (Figure 2) is provided to 2 variant architectures: base and large models, the base model is composed of 12 encoder layers and 12 multi-head attention heads. Whereas the large model comes with 24 encoder layers and multi-head attention heads.

We used a fine-tuned BERT model for the financial domain called FinBERT [25]. The model was pre-trained using Book corpus, Wikipedia, and Reuters TRC2 financial[4] , then fine-tuned using Financial Phrasebank [30]. Results of experiments on the Financial PhraseBank dataset classification task showed that FinBERT performed better than LSTM [31], ULMFit [32], LPS [30], HSC [33], FinSSLX [34] models.

The inputs of the classification component are the preprocessed social data which consists mainly of the tweet content and the social engagement information, the model takes as input the preprocessed tweet content and outputs a label as follows 0,1 or 2 respectively for positive, negative, or neutral sentiment. The output is then consolidated with selected tweet features including retweet, like, and replies count.

[4]https://trec.nist.gov/data/reuters/reuters.html

2) The time series prediction model

In this task, bitcoin price prediction is considered a time series problem. In fact, the sequential aspect of the data helps modeling it as a time series. We used an LSTM-based model to predict the price for the next day based on historical values of selected features for a window of 5 days. The features include both technical and social indicators Tables I and II.

We used Keras[5] to implement the price predictor, a model that consists of an LSTM layer with 50 neurons, a hyperbolic tangent activation function, a sigmoid as a recurrent activation, and an additional Dense layer fully connected to the LSTM layer. The model was trained on 100 epochs with a batch of 20 samples, using Adam [35] optimizer and Root Mean Square Error (RMSE) [36] loss function.

The price predictor takes as input the values of the selected features for the 5 previous days to predict the next day's value of a target feature, the high price feature in our case, and it's worth noting that this process can be performed for each feature. In order to assess the influence of each social feature on the price prediction, various experiments were conducted covering the combination of technical and social features whose results are highlighted

[5]https://keras.io

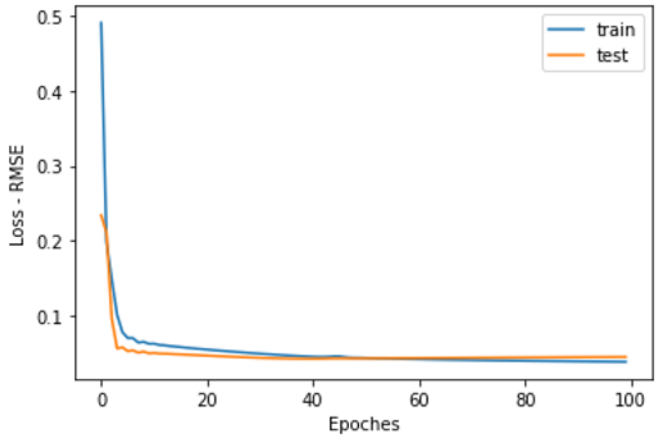

Figure 3. Forecasting model loss function by epochs

in table III. As for the training phase, the prepared dataset was divided for training and validation using a ratio of 80% for training and 20% for validation.

3) Evaluation

Time series prediction is a regression-type predictive modeling problem. There are several measurements that can fit our needs such as the RMSE measurement which is the standard deviation of the prediction errors calculated using equation 1, this measurement gives high weight to large errors which helps optimize models to get predictions with low error margin. In addition, to measure the level of correlation between our variables, the coefficient of determination R2 defined by equation 2 is commonly used in finance and helps decide which features are better predictors. The two measurements are often used to assess regression models quality; a low value of RMSE with a high R2 value generally indicates a good predictor.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2}{N}} \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (2)$$

y_i : Actual observations time series
$\tilde{y}_i$: Estimated observations
N : The number of data point
$\bar{y}_i$: The mean value of y_i

4. FINDINGS AND DISCUSSION

The predictability of bitcoin price by means of the selected features is a thought research question to answer. However, market and social data visual representation show a visual correlation between the selected features. In Figure 4, we plotted each feature during the selected period to conduct a visual assessment of features variations. Although bitcoin price open, high, low, and close variations are similar, sudden price movements can't be clearly explained through price features only. Furthermore, a sudden price drop in the month of May was accompanied by a high traded volume and tweet number which means that technical and social features hold information that may explain price variations.

We fine-tuned the LSTM predictor taking into consideration the lack and difficulty to aggregate training data based on the selected features. The model architecture was designed to avoid overfitting, Figure 3 shows the evolution of the loss function for both the training and validation data during the training process. The graph shows a quick convergence in the first 10 epochs, after that the gains from training become too minimal that we can stop early at epoch 50.

In order to assess the effect of the selected features on the predictability of bitcoin price, we trained our model with various data combinations; social, market, and a combination of social and market data. Table III shows that market or social data, alone, provides good results. However, the combination of all social and market features enhances notably our model performance. An $RMSE = 0.0368$ with $R^2 = 0.974$ indicates that the model is able to predict the high price feature with a low error margin; this can be performed with other price features or a combination of price features such as the mean bitcoin price per day.

The results of our study show that social features impact the price prediction to different degrees, the combination of each social feature with all the technical features reveals that likes count and total positive tweets provide good results compared to other social features, thus social engagement and overall positive sentiment are good bitcoin price predictors. Moreover, neutral tweets provide good performance and confirm results from [2] indicating that a large portion of tweets are neutral but convey information

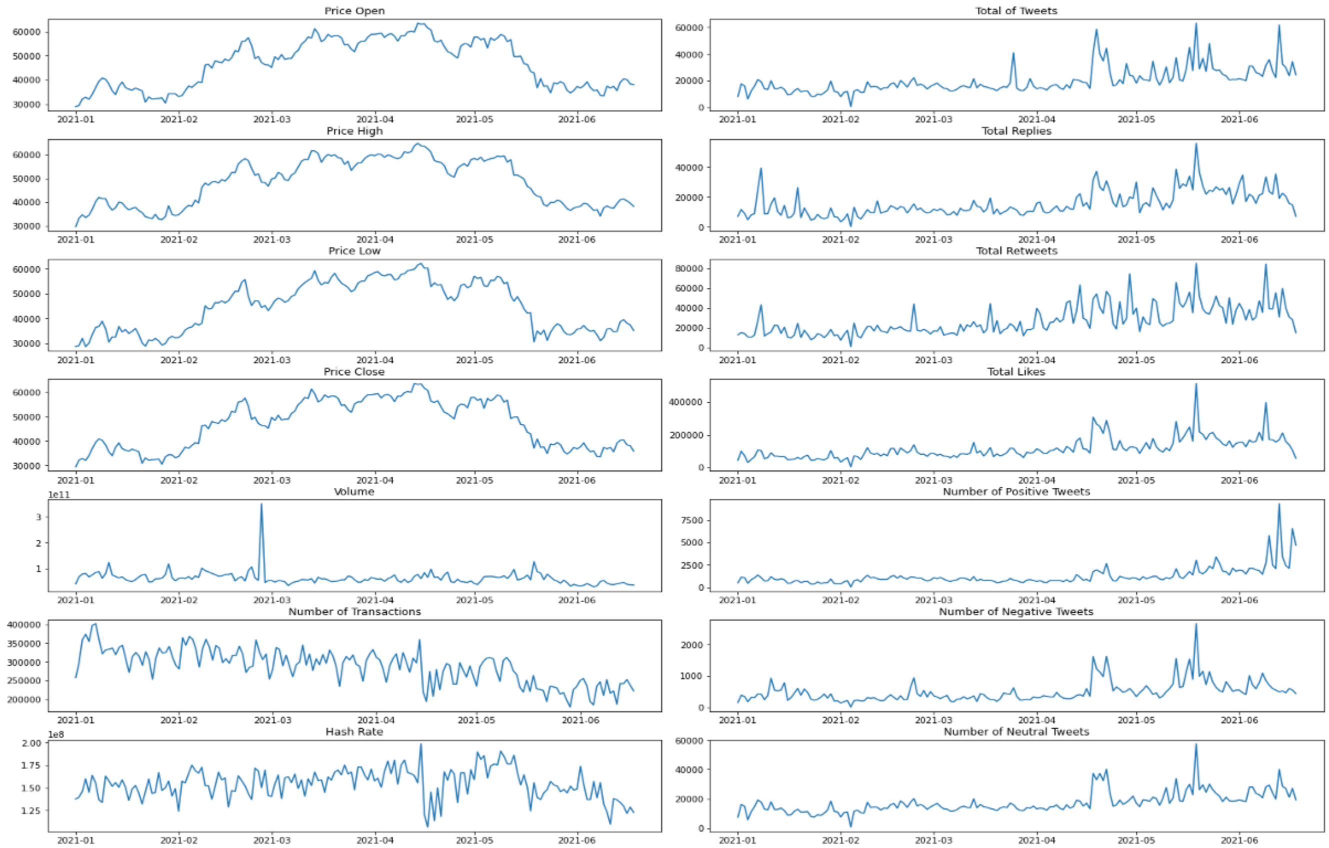

Figure 4. Social and technical factors variations during the period 1/1/21 to 18/6/21

Technical features	Social features	RMSE Train	RMSE Validation	R^2
ALL	ALL	0.0368	0.0435	0.974
ALL	-	0.0394	0.0544	0.96
-	ALL	0.0478	0.0565	0.956
ALL	Total Tweets	0.0403	0.0383	0.981
ALL	Replies Count	0.0405	0.0391	0.986
ALL	Retweets Count	0.0412	0.0353	0.985
ALL	Likes Count	0.0381	0.0556	0.958
ALL	Positive	0.0389	0.044	0.986
ALL	Negative	0.0434	0.0315	0.982
ALL	Neutral	0.0395	0.0395	0.984

Table III. FORECASTING MODEL PERFORMANCE WITH VARIOUS INPUT DATA

that may influence price fluctuations.

5. CONCLUSION

The interest in cryptocurrencies raises many research challenges, and predicting a cryptocurrency such as bitcoin's price has become a mainstream task in the SA research field. In our work, we proposed a multi-component system that ensures tweets classification using a transformer-based approach and an LSTM model for bitcoin price prediction using a combination of social and market features.

The sentiment classifier FinBERT relies on state-of-the-art results of BERT languages models, thus we have an accurate representation of bitcoin-related sentiment distribution from the gathered tweets. The social and technical information obtained was passed to the LSTM model after being processed to fit a time series problem, our choice is based on proven research results showcasing the success of recurrent neural networks with time series problems. The final system provided excellent results which may qualify it to be used as a baseline for future research systems.

As for the contribution of each social feature in the prediction of bitcoin price, our results show that social engagement is encoded in the number of likes and retweets, and the overall positive sentiment represented by the total positive and neutral tweets are good price predictors compared to other features. Moreover, the combination of all the selected social and technical features provides the best possible performance with our current system.

Although the proposed system provides good results, there are many rooms for improvement which open future research directions. The classification model can be further improved since the authors stated the limited ability of the model to deal with an implicit sentiment. Furthermore, the classified data can be enriched with other sources such as Reddit, news, forums, etc. this can provide a wide range of social information, a mixture of formal and informal language, and an accurate representation of public sentiment for bitcoin or other cryptocurrencies.

The social and technical features may increase in the future, thus predicting many steps ahead will raise the complexity of our model driving the need to review the current model architecture. Besides, it can be used to assess the contribution of each feature in the prediction of future cryptocurrency prices.

References

[1] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," p. 9, 2008.

[2] G. Cheuque Cerda and J. L. Reutter, "Bitcoin Price Prediction Through Opinion Mining," in *Companion Proceedings of The 2019 World Wide Web Conference*. San Francisco USA: ACM, May 2019, pp. 755–762.

[3] Z. Li, Y. Fan, B. Jiang, T. Lei, and W. Liu, "A survey on sentiment analysis and opinion mining for social multimedia," *Multimedia Tools and Applications*, vol. 78, no. 6, pp. 6939–6967, Mar. 2019. [Online]. Available: http://link.springer.com/10.1007/s11042-018-6445-z

[4] F. Fama, "SESSION TOPIC: STOCK MARKET PRICE BEHAVIOR," p. 35, 1965.

[5] B. Qian and K. Rasheed, "Stock market prediction with multiple classifiers," *Applied Intelligence*, vol. 26, no. 1, pp. 25–33, Jan. 2007. [Online]. Available: http://link.springer.com/10.1007/s10489-006-0001-7

[6] J. Bollen, H. Mao, and X.-J. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, Mar. 2011, arXiv: 1010.3003. [Online]. Available: http://arxiv.org/abs/1010.3003

[7] J. Abraham, D. Higdon, J. Nelson, and J. Ibarra, "Cryptocurrency Price Prediction Using Tweet Volumes and Sentiment Analysis," vol. 1, no. 3, p. 22, 2018.

[8] A. Sala, "Forecasting of the cryptocurrency market through social media sentiment analysis," p. 12, 2019.

[9] H. Guo, D. Zhang, S. Liu, L. Wang, and Y. Ding, "Bitcoin price forecasting: A perspective of underlying blockchain transactions," *Decision Support Systems*, vol. 151, p. 113650, Dec. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167923621001603

[10] K. Singpurwala, "Sentiment Analysis Trading Indicators," p. 11, 2021.

[11] M. McCoy and S. Rahimi, "Prediction of Highly Volatile Cryptocurrency Prices Using Social Media," *International Journal of Computational Intelligence and Applications*, vol. 19, no. 04, p. 2050025, Dec. 2020. [Online]. Available: https://www.worldscientific.com/doi/abs/10.1142/S146902682050025X

[12] B. El Haddaoui, R. Chiheb, R. Faizi, and A. El Afia, "Sentiment analysis: a review and framework foundations," *International Journal of Data Analysis Techniques and Strategies*, vol. 13, no. 4, p. 20, 2021.

[13] T.-M. Dulău and M. Dulău, "Cryptocurrency – Sentiment Analysis in Social Media," *Acta Marisiensis. Seria Technologica*, vol. 16, no. 2, pp. 1–6, Dec. 2019. [Online]. Available: https://content.sciendo.com/view/journals/amset/16/2/article-p1.xml

[14] O. Kraaijeveld and J. De Smedt, "The predictive power of public Twitter sentiment for forecasting cryptocurrency prices," *Journal of International Financial Markets, Institutions and Money*, vol. 65, p. 101188, Mar. 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S104244312030072X

[15] K. Wołk, "Advanced social media sentiment analysis for short-term cryptocurrency price prediction," *Expert Systems*, vol. 37, no. 2, Apr. 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/exsy.12493

[16] C. Lamon, E. Nielsen, and E. Redondo, "Cryptocurrency Price Prediction Using News and Social Media Sentiment," p. 6, 2017.

[17] S. Colianni, S. Rosales, and M. Signorotti, "Algorithmic Trading of Cryptocurrency Based on Twitter Sentiment Analysis," p. 5, 2015.

[18] C. J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," p. 10, 2014.

[19] T. Wilson, P. Hoffmann, S. Somasundaran, J. Kessler, J. Wiebe, Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan, "OpinionFinder: a system for subjectivity analysis," in *Proceedings of HLT/EMNLP on Interactive Demonstrations -*. Vancouver, British Columbia, Canada: Association for Computational Linguistics, 2005, pp. 34–35.

[20] H.-M. Kim, G.-W. Bock, and G. Lee, "Predicting Ethereum prices with machine learning based on Blockchain information," *Expert Systems with Applications*, vol. 184, p. 115480, Dec. 2021. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0957417421008915

[21] Z. H. Kilimci, "Sentiment Analysis Based Direction Prediction in Bitcoin using Deep Learning Algorithms and Word Embedding Models," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 8, no. 2, pp. 60–65, Jun. 2020. [Online]. Available: https://ijisae.org/IJISAE/article/view/1062/616

[22] I. E. Livieris, N. Kiriakidou, S. Stavroyiannis, and P. Pintelas, "An Advanced CNN-LSTM Model for Cryptocurrency Forecasting," *Electronics*, vol. 10, no. 3, p. 287, Jan. 2021. [Online]. Available: https://www.mdpi.com/2079-9292/10/3/287

[23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," *arXiv:1706.03762 [cs]*, Dec. 2017, arXiv: 1706.03762. [Online]. Available: http://arxiv.org/abs/1706.03762

[24] K. Mishev, A. Gjorgjevikj, I. Vodenska, L. T. Chitkushev, and D. Trajanov, "Evaluation of Sentiment Analysis in Finance: From Lexicons to Transformers," *IEEE Access*, vol. 8, pp. 131 662–131 682, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9142175/

[25] D. Araci, "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models," *arXiv:1908.10063 [cs]*, Aug. 2019, arXiv: 1908.10063. [Online]. Available: http://arxiv.org/abs/1908.10063

[26] X. Huang, W. Zhang, X. Tang, M. Zhang, J. Surbiryala, V. Iosifidis, Z. Liu, and J. Zhang, "LSTM Based Sentiment Analysis for Cryptocurrency Prediction," *arXiv:2103.14804 [cs]*, Oct. 2021, arXiv: 2103.14804. [Online]. Available: http://arxiv.org/abs/2103.14804

[27] S. McNally, J. Roche, and S. Caton, "Predicting the Price of Bitcoin Using Machine Learning," in *2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP)*. Cambridge: IEEE, Mar. 2018, pp. 339–343.

[28] R. Mohammdi Farsani and E. Pazouki, "A Transformer Self-attention Model for Time Series Forecasting," *Journal of Electrical and Computer Engineering Innovations (JECEI)*, no. Online First, Dec. 2020. [Online]. Available: https://doi.org/10.22061/jecei.2020.7426.391

[29] F. Dama and C. Sinoquet, "Time Series Analysis and Modeling to Forecast: a Survey," *arXiv:2104.00164 [cs]*, Sep. 2021, arXiv: 2104.00164. [Online]. Available: http://arxiv.org/abs/2104.00164

[30] P. Malo, A. Sinha, P. Takala, P. Korhonen, and J. Wallenius, "Good Debt or Bad Debt: Detecting Semantic Orientations in Economic Texts," *arXiv:1307.5336 [cs, q-fin]*, Jul. 2013, arXiv: 1307.5336. [Online]. Available: http://arxiv.org/abs/1307.5336

[31] S. Hochreiter, "The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 06, no. 02, pp. 107–116, Apr. 1998. [Online]. Available: https://www.worldscientific.com/doi/abs/10.1142/S0218488598000094

[32] J. Howard and S. Ruder, "Universal Language Model Fine-tuning for Text Classification," *arXiv:1801.06146 [cs, stat]*, May 2018, arXiv: 1801.06146. [Online]. Available: http://arxiv.org/abs/1801.06146

[33] S. Krishnamoorthy, "Sentiment Analysis of Financial News Articles using Performance Indicators," *Knowledge and Information Systems*, vol. 56, no. 2, pp. 373–394, Aug. 2018, arXiv: 1811.11008. [Online]. Available: http://arxiv.org/abs/1811.11008

[34] M. Maia, A. Freitas, and S. Handschuh, "FinSSLx: A Sentiment Analysis Model for the Financial Domain Using Text Simplification," in *2018 IEEE 12th International Conference on Semantic Computing (ICSC)*. Laguna Hills, CA: IEEE, Jan. 2018, pp. 318–319.

[35] T. Dozat, "INCORPORATING NESTEROV MOMENTUM INTO ADAM," p. 4, 2016.

[36] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247–1250, Jun. 2014. [Online]. Available: https://gmd.copernicus.org/articles/7/1247/2014/

Bousselham EL HADDAOUI is a Phd student specialized in Opinion Mining and Sentiment Analysis, Independent IT Advisor in digital communication and event management software. He has an engineering degree in software engineering from ENSIAS, Mohamed V University in Rabat Morocco, 2012.

Raddouane CHIHEB is a professor of Higher Education, Head of the Computer Science and Decision Support Department, PhD in Applied Mathematics from Jean Monnet University of Saint-E Etienne, 1998. Specialized Master in Computer Science from Insa in Lyon, France, 2001, and the President of the Moroccan Association for Value Analysis.

Rdouan FAIZI is a Full Professor at National School of Computer Science and Systems Analysis (ENSIAS). He obtained his PhD in English Language and Literature from Mohammed V Agdal University, 2002. Research areas of interest are Linguistics, E-learning, Business English Skills, TOEIC Preparation, Scientific Communication.

Abdellatif EL AFIA is a Full Professor, PhD in Operational Research from the University of Sherbrooke. 1999, and Coordinator of the Artificial Intelligence Engineering branch (2IA).