

Money Often Costs Too Much: A Study to Investigate The Effect Of Twitter Sentiment On Bitcoin Price Fluctuation

Mitul Verma
School Of Computing
Dublin City University
Dublin, Ireland
Email:mitul.verma3@mail.dcu.ie

Pritish Sharma
School Of Computing
Dublin City University
Dublin, Ireland
Email:prish.sharma3@mail.dcu.ie

Abstract—Introduced in 2009, Bitcoin has demonstrated a huge potential as the world's first digital currency and has been widely used as a financial investment. Our research aims to uncover the relationship between Bitcoin prices and people's sentiments about Bitcoin on social media. Among various social media platforms, micro-blogging is one of the most popular. Millions of people use micro-blogging platforms to exchange ideas, broadcast views, and to provide opinions on different topics related to politics, culture, science, and technology. This makes them a potentially rich source of data for sentiment analysis. Therefore we chose one of the busiest micro-blogging platforms, Twitter, to perform sentiment analysis on Bitcoin. We used ELMo embedding model to convert Bitcoin-related tweets into a vector form and SVM classifier to divide the tweets into three sentiment categories - positive, negative, and neutral. We then used the sentiment data to find its relation with Bitcoin price fluctuation using the linear mixed model.

1. Introduction

Micro-blogging, as the name suggests, refers to a broadcast medium where content consists of short sentences, individual images, or video links [1]. Twitter, Tumblr, and Pinterest are notable micro-blogging platforms. Twitter is the most prominent micro-blogging platform with 166 million daily active users as of Q1 2020 [2]. Users can post micro-posts or tweets which can be up to 280 characters long and attach images, videos, GIFs, and links to websites with tweets. Twitter provides a convenient way for millions around the world to broadcast views globally. Furthermore, tweets can be marked with hashtags and these hashtags can be used to search for tweets on a specific topic. On Twitter, 166 million daily active users contribute 500 million tweets per day on a variety of topics [3]. We have used a dataset consisting of tweets related to Bitcoin as a source for sentiment analysis.

Bitcoin was introduced in 2008 as a decentralized digital currency that uses peer to peer network for transactions, eliminating the need for a central authority like a central bank [4]. Initially, Bitcoin was primarily adopted by

black markets such as Silk Road [5] which exclusively accepted Bitcoins as payment but over time Bitcoin has gained wider acceptance by online merchants and vendors [6]. Venture capitalists have invested in various Bitcoin-based startups since 2012 [7], [8] and in December 2017, the Chicago Board Options Exchange and Chicago Mercantile Exchange started trading Bitcoin futures [9], [10]. All these developments popularised Bitcoin as an investment option which resulted in the spectacular increase in its price in 2017, climbing to an all-time high of USD 19,783.06 on 17 December 2017 [11]. This invited the curiosity of the general public and further popularised the idea of alternative currencies. Bitcoin has not only revolutionized the concept of digital currencies but has also cultivated the seeds for tech development based on blockchain such as smart contracts [12], peer to peer energy trading [13], and blockchain-based supply chain management [14]. As of August 1, 2020, the number of cryptocurrencies exceeds 2000 and the market capitalization of cryptocurrencies stands at over 300 Billion USD [15]. Before one invests in the crypto trading options, it is very important to understand the nature and behavior of this new and emerging market. Various factors affect the price of the cryptocurrency and understanding them is crucial to understanding the crypto market. This paper is solely dedicated to finding a correlation, if exists, between Bitcoin price and people's sentiments on Twitter and how closely they are related.

We start by first defining our hypothesis which lays the foundation for our research followed by the research questions that we aim to answer. In the *Literature Review* section, we examined various techniques used to represent text into word embedding, methodologies to perform sentiment analysis, enhancements in the field, and drawbacks and advantages of each approach. We also reviewed the literature on the Bitcoin price prediction and have summarised the methodology and techniques by researchers to identify the impact of various factors on Bitcoin price fluctuation. Furthermore, we explored the research papers on the challenges and complexities involved in performing sentiment analysis on a multilingual text.

Once equipped with adequate knowledge of the litera-

ture, we devised our methodology which is described in the *Methodology and Results* section. In the first subsection, we outline the data collection and transformation process for our tweet dataset followed by data preparation which involved manually labeling 35,000 tweets into sentiment categories - positive, negative, or neutral and converting the tweets into word embedding. We implemented various classification algorithms on our labeled data and to identify the best algorithm, we performed a resampling procedure and a statistical test. We classified the remaining tweets into sentiment categories using the selected algorithm. Finally, we used a linear mixed model to determine the relationship between Bitcoin prices and sentiments. In the end, we discuss the results and suggest future work to further advance the research.

2. Hypothesis

Inspired by the work and enhancements in the field of sentiment analysis [16], [17] and the future potential of Bitcoin as the world's first digital currency and new investment option, we tried to understand the effect of people's sentiments on Bitcoin price. To understand this effect we bind our whole research into the following hypothesis - Ho: There is no relationship between Bitcoin price and Twitter sentiments.

3. Research Question

To reject our null hypothesis we formulate some questions and try to answer these questions through our methodology. These questions form the base of our research. Our first research question was i) Are tweets on Bitcoins a good source for sentiment analysis on Bitcoin. ii) Can a model be created which can accurately classify tweets into different sentiment categories. iii) Does there exist a correlation between Bitcoin price and people's sentiments on Twitter. To answer the first research question we manually read and labeled a sample of tweets to check if tweets can be categorized into positive, negative, or neutral sentiments. To answer the second research question we used ELMo embedding to convert our text data into vectors of embedding. We trained different machine learning classifiers and selected the one with the highest accuracy to classify our data into three categories. We further performed significance tests to support our claim. For our last research question, we use a linear mixed model to find the correlation between Bitcoin prices and sentiments.

4. Flowchart

Figure 1 shows the diagrammatic representation of workflow or step-wise design of the whole research work. The rectangular boxes show the different processes, quadrilateral represents the significance test and arrows show the sequential order of processes.

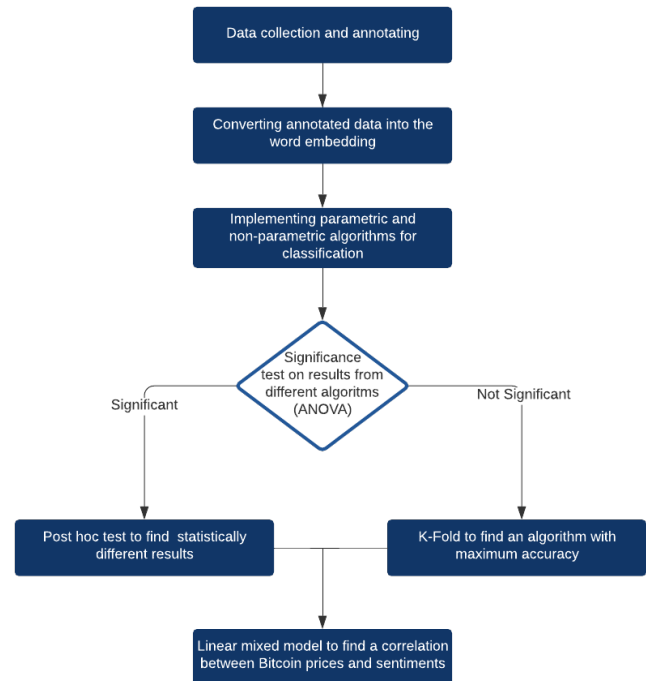


Figure 1. Flow of research

5. Literature Review

The literature review is divided into three sections. The first section is on sentiment analysis. It contains research on how sentiment analysis is performed, the different methodologies used by different researchers, results, and enhancements with time. The second section is on Bitcoin price fluctuation, factors that affect Bitcoin price, and various models used by researchers to find correlations and predict Bitcoin price. In the third section, we look at the challenges with sentiment analysis on multilingual texts and describe why we chose only English for sentiment analysis.

5.1. Sentiment Analysis

Opinion mining is one of the key factors in understanding people's perspectives and creating new strategies in many key areas. Much of the work on sentiment analysis involves classifying documents by their overall sentiments. Tweets are classified into positive and negative by using polarity scores [18]. In this paper, the researchers have classified tweets based on polarity scores with an accuracy score of 80.6 percent. The results are used to understand the effect on the stock price. The author found that the approach was good with respect to time complexity and accuracy but did not consider other aspects of a sentence e.g. syntactic or syntax and semantic, and arbitrary weights were assigned to sentences using polarity score. Further developments in the field of sentiment analysis included models like unigram, feature-based, and tree kernel-based models [19]. The results

show improvement in sentiment scores. This approach was better than the one discussed before [18] as the prior priority of words and their part-of-speech tags both were used to create a feature in this approach. While semantic analysis, topic modeling, and parsing were not considered during feature creation. Later, the development of WordNet 3.0, an enhanced lexical analysis for opinion mining, produced better results than previously used techniques [20]. The work presents the use of enhanced lexical resources designed for opinion mining applications and for supporting sentiment classification. The results showed significant improvement. 19.4 and 21.96 percent relative improvement in classifying positive and negative sentiment as well as improvement from other baseline models. WordNet is a rich set of vocabulary created using thousands of words, their meaning, part of speech, and degree of positivity and negativity, ranging from 0 to 1. WordNet is widely used by researchers for text classification. But WordNet does not consider syntax and semantics which is a shortcoming of this work [21]. A significant work in natural language processing emerges through the development of linguistic analysis [19]. In linguistic analysis, not only the polarity of words is considered but the whole sentence is considered, adding more meaning and weight to the analysis to perform text classification. The classifier created by using the corpus generated through linguistic analysis performs better than the rest of the baseline models. The performance of the model is measured using the F Score given by formula [19]. The B-Gram model was found to outperform the N-gram model and that attaching negation words provides very high accuracy. Thus the classifier generates a very accurate result. Although the accuracy was found to be better than all previous researches but it does not improve beyond a certain point. The new approach *phrase-level sentiment analysis* as discussed in [22] changes the overall perspective on classifying text. All previous work as discussed in [18], [19], [20] only includes identifying positive and negative reviews. But to seek an answer for complicated problems like a) mining product reviews b) multi-perspective questioning and answering and summarization, opinion-oriented information extraction, a better approach was needed. To answer all these questions accurately we required sentence-level or phrase-level sentiment analysis. The researchers used the phrase-level sentiment analysis to answer all these questions. The first step is to determine whether an expression is neutral or polar followed by disambiguating the polarity of the polar expression. Using this approach, the system automatically identifies contextual polarity even for large datasets with improved accuracy. The result is also better than the baseline models. Even for the simplest classifier, the accuracy was 48 percent higher than the classifiers used previously.

Finally, the advancement in opinion mining led to the development of a content-based approach. Content-based approaches are finest of all but are computationally expensive [21]. Many embedding models are aimed to provide transfer learning to a wide variety of natural language processing tasks after being trained on huge corpora to learn deep semantic representation. One of the popular pre-trained

models of all is the ELMo(ELMO) [21]. ELMo is a pre-trained machine learning model that can encode sentences into deeply contextualized embeddings. The results show that such models have outperformed previous state-of-art approaches such as traditional word embeddings for many natural language processing tasks [18], [19], [20], [23], [22]. Such NLP tasks include calculating semantic similarity and relatedness, text classification, etc. Such advancement has led to the execution of complicated tasks such as calculating document relatedness in case of a research paper recommender system [21]. The result generated shows that ELMo outperformed other baselines models but was outperformed by USE, BERT, And SciBERT. The possible reason was that ELMo was trained on corpora that contain news crawls and natural language inference while the latter are trained on the technical and scientific texts.

5.2. Bitcoin Analysis

Launched in 2008, Bitcoin has gained immense popularity owing to its spectacular adoption and stunning price increase from 2013 onwards. Such popularity also opened various trading and investment options with Bitcoin such as Bitcoin futures [9]. The sentiment is always considered an important driver in Bitcoin price fluctuation especially during the duration of excessive volatility [24]. The researchers used the Autoregressive(AR) Model to show the relationship between sentiments and Bitcoin prices. They were successful in capturing a significant impact of positive sentiments on the prices. They also realized that positive sentiments have more impact than negative sentiments. The results only present a marginal explanatory value of price fluctuation of Bitcoin due to sentiments. The results were as expected because only one source of sentiment was used. In the next paper, the researchers used a time-series model (Vector error correction model) to analyse and study the relationship between various economic indicators, tweets, and technological factors [25]. Sentiment analysis was performed on a daily basis using a support vector machine(SVMs). The results generated show that Bitcoin prices are positively related to positive sentiments, Wikipedia searches, and hash rate (a measure of mining difficulty), while the USD and EURO exchange rates have a negative impact on Bitcoin. In the long-run analysis, the Bitcoin price and poor S&P 500 performance have positive and negative impact respectively. The paper shows significant findings and scope for future research. The dataset used by researchers was not very big and a vector autoregressive model might be better instead of OLS to study the short-term dynamics of price and sentiment analysis can be used to predict long-term prices by using appropriate algorithmic processes. Investors were also interested in predicting the price of cryptocurrencies other than Bitcoin (also called altcoins) as they offer equal investing potential. Researchers also study the pattern of another cryptocurrency (Ether) and compare its behavior with Bitcoin [26]. In this paper, researchers used Twitter data and Google trends data to understand the price fluctuations of the two most popular cryptocurrencies, Bitcoin, and Ether.

many different languages which included Chinese, Spanish, French, etc. After cleaning the data, we were left with 190,000 records for the whole year. The cleaning process involved removing special characters, digits, punctuation marks, and non-English sentences. We also removed stop words after labeling the sentences into three categories positive, negative, and neutral.

6.2. Data Preparation

To prepare our data we manually labeled 35,000 tweets into positive, negative, and neutral sentiments. To evaluate the validity of our labeled dataset we calculated the Kappa coefficient [33]. The equation for the Kappa coefficient (Eq.1). Where P_o is the relative observed agreement among raters. P_e is the hypothetical probability of chance agreement. Kappa coefficient, also known as Cohen's Kappa coefficient, is a statistical method used to measure the inter-rater reliability for categorical data. The Kappa coefficient value for our data was 0.83 which proves that we had almost similarly labeled our data individually into positive, negative, and neutral and there exists no conflict of opinion between the authors of this paper. Next, we used ELMo embedding to convert the sentences into word embedding. Word Embedding is the numerical representation of a string of words. In word embedding, similar strings are represented by similar embedding. Figure 2 show diagrammatic representation of word embedding [34]. ELMo is a bi-directional LSTM language model used to compute contextualized character-based word representation [21]. It has two layers stacked together. Each layer has two pass one forward and backward. In the bi-directional language model, both the left and right contexts of the target word are used. It means that the target word can be predicted either by using the left context or the right context [35]. Diagrammatic representation of ELMo embedding is shown in figure 3 [34]. While most of the embeddings only capture the frequency of words to give them the weight and hence are unable to capture the meaning of words, ELMo embedding captures both the syntax and semantics [36]. We used the following ELMo tensorflow configuration trained on one billion word benchmark shown in table 4. Further, this module supports text in both forms, string as well as tokenized. The rest of the ELMo parameters are used at default configuration with fixed embedding at each LSTM layer, 3 layers learnable aggregation, and a fixed mean-pooled vector representation of the input. We split our dataset into a trained and test dataset with a ratio of 70:30 which is a standard ratio to follow.

$$k = \frac{p_0 - p_e}{1 - p_e} \quad (1)$$

6.3. Model Training And Implementation Of Algorithms

Now that we had the labeled data for 35,000 records and ELMo representation of all the tweets. We needed to find the

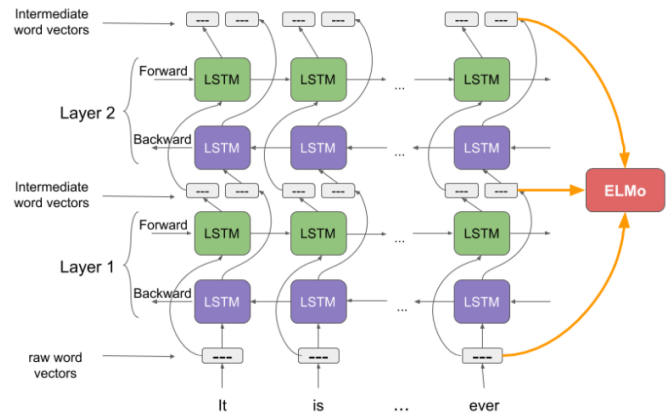


Figure 3. Diagrammatic representation of ELMo embedding

Tensorflow	==	1.15
Tensorflow_hub	>=	0.6.0
Tensorflow_text	==	1.15

Figure 4. Tensorflow Configuration

best classifier to classify the remaining tweets into positive, negative, and neutral categories. For training and testing our algorithms, we used an equal number of records from each of the sentiment categories i.e. the same number of positive, negative, and neutral records. We implemented one parametric [37] and three non-parametric machine learning models [38]. For the non-parametric models, we used three classifier algorithms. They are, i) KNN neighbours [39], ii) Decision Tree [40] and iii) SVM [41]. For the parametric model, we used only one classifier algorithm, Naïve Bayes [42].

We used K-fold technique to find the algorithm with maximum accuracy [43]. The results of the four algorithms, where K-fold is equal to 10, are shown in a highlighted table 5. Red-green diverging color gradient shows how significantly different the results are for each algorithm at each fold. The table also shows the average results for all algorithms. From the results, we can conclude that SVM is a better classifier algorithm than the other algorithms under

K-Fold	Naïve Bayes	Decision Tree	KNN Classifier	SVM Classifier
K=1	0.63	0.76	0.88	0.94
K=2	0.43	0.88	0.87	0.93
K=3	0.43	0.88	0.86	0.91
K=4	0.44	0.8	0.78	0.87
K=5	0.53	0.81	0.81	0.9
K=6	0.74	0.83	0.88	0.92
K=7	0.76	0.89	0.89	0.93
K=8	0.73	0.86	0.86	0.9
K=9	0.72	0.86	0.88	0.9
K=10	0.73	0.87	0.84	0.92
Average	0.614	0.844	0.855	0.912

Figure 5. K-Fold Result

	SS	df	MS	F	PR(>F)
SSB	0.52	3	0.15	29.29	0.003
SSW	0.21	36	0.005	-	-

TABLE 1. ANOVA ANALYSIS

A	B	mean(A)	p-tukey	hedges
Decision Tree	KNN	0.85	0.9	-0.12
Decision Tree	Naive Bayes	0.85	0.0	2.86
Decision Tree	SVM Classifier	0.85	0.19	-0.8
KNN	Naive Bayes	0.86	0.001	2.98
KNN	SVM Classifier	0.86	0.32	-0.72
Naive Bayes	SVM Classifier	0.61	0.001	-3.71

TABLE 2. TUKEY HSD

consideration as its average accuracy is comparatively better. To further authenticate this result we perform a significance test [44]. We use ANOVA followed by post-hoc comparison test [45]. To conduct ANOVA analysis, we formulate the following hypothesis "H0: There is no significant difference between machine learning algorithms". Once we had rejected the null hypothesis we also performed a post-hoc comparison test to identify the machine algorithm that is significantly different from all other algorithms. For post-hoc comparison we use Tukey HSD [46]. Table 1 shows the ANOVA result.

From table 1, the p-value is less than 0.05. Where 0.05 is 5% significance level. So, we reject our null hypothesis which states that "H0:-There is no significant difference among algorithms" and accept the alternate hypothesis. Further, we use a post-hoc test to find the algorithms which are significantly different from each other. Table 2 shows the result of Tukey HSD test.

From the result of the post-hoc test table 2, we conclude that the accuracy of the non-parametric models is significantly different from that of a parametric model. In other words, the non-parametric models are better than the parametric model for our research work. This is because there exists a bias in data. The dimension of words from ELMo embedding is 1024. While the number of records used to train and test algorithms is 27,000. This introduced bias in the train and test data. So, the non-parametric models produced better results because they have high variance [47] that balances the bias in our data. Further analysis from the post-hoc test shows that there is no significant difference among different non-parametric models used. We chose SVM classifier over other non-parametric models because after performing K-fold cross validation at 10-fold the mean accuracy of SVM is 0.92 approx figure 5, which was the maximum of all models. So, we selected the combination of ELMo+SVM to label our complete dataset.

6.4. Finding Correlation Between Bitcoin And Twitter Sentiments

In this last part, we use the linear mixed model to understand the effect of Twitter sentiments on Bitcoin price. Linear mixed models are used when we need to find a correlation between the dependent and independent variables.

$$y = X\beta + Zu + \epsilon$$

y is a $N \times 1$ column vector, the outcome variable.

X is a $N \times p$ matrix of the p predictor variables.

β is a $p \times 1$ column vector of the fixed-effects regression coefficients (the β s).

Z is the $N \times qJ$ design matrix for the q random effects and J groups.

u is a $qJ \times 1$ vector of q random effects for J groups.

ϵ is a $N \times 1$ column vector of the residuals.

Figure 6. Linear mixed model

```

=====
Model:                MixedLM  Dependent Variable: Difference
No. Observations:    365      Method:                REML
No. Groups:          267      Scale:                 26731.5210
Min. group size:     1        Likelihood:            -2651.4797
Max. group size:     5        Converged:              Yes
Mean group size:     1.4
=====

```

	Coef.	Std.Err.	z	P> z	[0.025	0.975]
Intercept	-31.185	58.767	-0.531	0.596	-146.366	83.995
Positive	0.033	0.510	0.066	0.948	-0.967	1.034
Negative	-0.437	0.736	-0.593	0.553	-1.879	1.006
Neutral	0.405	0.294	1.376	0.169	-0.172	0.981
Group Var	151392.947	194.843				

Figure 7. Linear mixed model regression analysis at 5% significant level

Linear models are also useful when we need to consider both fixed and random effects [48]. The equation is shown in figure 6. In our model, we consider per day count of positive, negative, and neutral sentiments as fixed effect and volume (total no of positive, negative, and neutral) as a random effect. While the dependent variable was difference between per day opening and closing price. The model is tested on a different level of significance. Figure 7 show result at 5% significance level, figure 8 at 1% significance level, and figure 9 at 10% significance level.

From the results, we can see our p-value is more than the level of significance at all the significance level under

```

=====
Mixed Linear Model Regression Results
=====
Model:                MixedLM  Dependent Variable: Difference
No. Observations:    365      Method:                REML
No. Groups:          267      Scale:                 26731.5210
Min. group size:     1        Likelihood:            -2651.4797
Max. group size:     5        Converged:              Yes
Mean group size:     1.4
=====

```

	Coef.	Std.Err.	z	P> z	[0.005	0.995]
Intercept	-31.185	58.767	-0.531	0.596	-182.558	120.188
Positive	0.033	0.510	0.066	0.948	-1.281	1.348
Negative	-0.437	0.736	-0.593	0.553	-2.333	1.459
Neutral	0.405	0.294	1.376	0.169	-0.353	1.162
Group Var	151392.947	194.843				

Figure 8. Linear mixed model regression analysis at 1% significant level

Mixed Linear Model Regression Results						
=====						
Model:	MixedLM	Dependent Variable:		Difference		
No. Observations:	365	Method:	REML			
No. Groups:	267	Scale:	26731.5210			
Min. group size:	1	Likelihood:	-2651.4797			
Max. group size:	5	Converged:	Yes			
Mean group size:	1.4					

	Coef.	Std.Err.	z	P> z	[0.05	0.95]

Intercept	-31.185	58.767	-0.531	0.596	-127.848	65.477
Positive	0.033	0.510	0.066	0.948	-0.806	0.873
Negative	-0.437	0.736	-0.593	0.553	-1.647	0.774
Neutral	0.405	0.294	1.376	0.169	-0.079	0.888
Group Var	151392.947	194.843				
=====						

Figure 9. Linear mixed model regression analysis at 10% significant level

consideration. Hence, we do not have enough evidence to reject the null hypothesis. Therefore we accept the null hypothesis that there exists no relationship between sentiments and Bitcoin price.

7. Conclusion and Future work

From the results, we can conclude that Bitcoin prices are not affected by people's sentiments on Twitter. So, it might be possible that other factors such as technology, government policies, institutional investors, new ventures, partnerships, news on cryptocurrency, and new crypto technologies may affect the price of Bitcoin. The result generated also scratches on the possibility that Bitcoin price might affect people's opinions/sentiments. In other words, the price fluctuation of Bitcoin controls the emotion of people on Twitter. So, whatever people tweet on Bitcoin, either positive or negative, is because of Bitcoin price fluctuation. For example, someone holding a share of Bitcoin will express gratitude if the price rises while others may show sadness and anger. So, the results show that tweets on Bitcoin are nothing more than just the aftermath of Bitcoin price fluctuations.

The above research is restricted to only one language(English). So, it can be extended further to multiple languages. For word-embedding, a new advanced model can be used such as Bert. Instead of tweets, news headlines can be used as a source of sentiments as well as statements by prominent personalities. The period used is limited to a single year when the price of Bitcoin was highly volatile. The research can be extended to 2-3 years of data. The above work is limited to only a single cryptocurrency. It is possible that different cryptocurrencies behave differently in regards to people's sentiments. Also, the above work does not talk about the relationship between cryptocurrencies such as how the price of one affects the other or why Bitcoin is more volatile than all other cryptocurrencies. Weights can be assigned to each sentiment instead of labels depending on the type of context, the person who released the statement, and likes on a comment. Thus per day total weight can be used instead of per day total counts of different tweets.

Acknowledgment

We would like to thank Prof. Martin Crane and other staff members of Dublin city University for their valuable feedback and suggestions. There is no external funding required in this project. The authors are thank full to everyone who directly or indirectly involved in providing the support to our research project.

References

- [1] A. Kaplan and M. Haenlein, "The early bird catches the news: Nine things you should know about micro-blogging," *Business Horizons*, vol. 54, pp. 105–113, 03 2011.
- [2] "Quarterly results," Apr 2020. [Online]. Available: <https://investor.twitterinc.com/financial-information/quarterly-results/>
- [3] D. Sayce, "The number of tweets per day in 2020," Jul 2020. [Online]. Available: <https://www.dsayce.com/social-media/tweets-day/>
- [4] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," Oct 2008. [Online]. Available: <https://bitcoin.org/bitcoin.pdf>
- [5] R. Böhme, N. Christin, B. Edelman, and T. Moore, "Bitcoin: Economics, technology, and governance," *Journal of economic Perspectives*, vol. 29, no. 2, pp. 213–38, 2015.
- [6] A. Cuthbertson, "Bitcoin is now accepted at starbucks, whole foods and 100s of other shops," May 2019. [Online]. Available: <https://www.independent.co.uk/life-style/gadgets-and-tech/news/bitcoin-stores-spend-where-starbucks-whole-foods-crypto-a8913366.html>
- [7] T. Simonite, "Bitcoin millionaires become investing angels," Apr 2020. [Online]. Available: <https://www.technologyreview.com/2013/06/12/15919/bitcoin-millionaires-become-investing-angels/>
- [8] "Ten-hut! bitcoin recruits snap to," Dec 2014. [Online]. Available: <https://www.wsj.com/articles/DJFVW00120141201eac1ag1qi>
- [9] CBS, "Chicago mercantile exchange jumps into bitcoin futures," Dec 2017. [Online]. Available: <https://www.cbsnews.com/news/chicago-mercantile-exchange-jumps-into-bitcoi-futures/>
- [10] CBS, "Bitcoin futures surge in first day of trading," Dec 2017. [Online]. Available: <https://www.cbsnews.com/news/bitcoin-futures-price-surge-in-first-day-of-trading/>
- [11] D. Z. Morris, "Bitcoin hits a new record high, but stops short of 20,000 usd," Dec 2017. [Online]. Available: <http://fortune.com/2017/12/17/bitcoin-record-high-short-of-20000>
- [12] A. Kosba, A. Miller, E. Shi, Z. Wen, and C. Papamanthou, "Hawk: The blockchain model of cryptography and privacy-preserving smart contracts," in *2016 IEEE symposium on security and privacy (SP)*. IEEE, 2016, pp. 839–858.
- [13] M. Corkery and N. Popper, "From farm to blockchain: Walmart tracks its lettuce," Sep 2018. [Online]. Available: <https://www.nytimes.com/2018/09/24/business/walmart-blockchain-lettuce.html>
- [14] M. Andoni, V. Robu, D. Flynn, S. Abram, D. Geach, D. Jenkins, P. McCallum, and A. Peacock, "Blockchain technology in the energy sector: A systematic review of challenges and opportunities," *Renewable and Sustainable Energy Reviews*, vol. 100, pp. 143–174, 2019.
- [15] "Global charts." [Online]. Available: <https://coinmarketcap.com/charts/>
- [16] S. Venkatakrisnan, A. Kaushik, and J. K. Verma, "Sentiment analysis on google play store data using deep learning," in *Applications of Machine Learning*. Springer, 2020, pp. 15–30.

- [17] A. Kaushik and S. Naithani, "A comprehensive study of text mining approach," *International Journal of Computer Science and Network Security (IJCSNS)*, vol. 16, no. 2, p. 69, 2016.
- [18] R. K. Bakshi, N. Kaur, R. Kaur, and G. Kaur, "Opinion mining and sentiment analysis," in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*. IEEE, 2016, pp. 452–455.
- [19] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. J. Passonneau, "Sentiment analysis of twitter data," in *Proceedings of the workshop on language in social media (LSM 2011)*, 2011, pp. 30–38.
- [20] S. Baccianella, A. Esuli, and F. Sebastiani, "Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining." in *Lrec*, vol. 10, no. 2010, 2010, pp. 2200–2204.
- [21] H. A. M. Hassan, G. Sansonetti, F. Gasparetti, A. Micarelli, and J. Beel, "Bert, elmo, use and infersent sentence encoders: The panacea for research-paper recommendation?" in *RecSys (Late-Breaking Results)*, 2019, pp. 6–10.
- [22] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," in *Proceedings of human language technology conference and conference on empirical methods in natural language processing*, 2005, pp. 347–354.
- [23] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining." in *LREC*, vol. 10, no. 2010, 2010, pp. 1320–1326.
- [24] J. Bukovina, M. Marticek *et al.*, "Sentiment and bitcoin volatility," *MENDELU Working Papers in Business and Economics*, vol. 58, 2016.
- [25] I. Georgoula, D. Pournarakis, C. Bilanakos, D. Sotiropoulos, and G. M. Giaglis, "Using time-series and sentiment analysis to detect the determinants of bitcoin prices," *Available at SSRN 2607167*, 2015.
- [26] J. Abraham, D. Higdon, J. Nelson, and J. Ibarra, "Cryptocurrency price prediction using tweet volumes and sentiment analysis," *SMU Data Science Review*, vol. 1, no. 3, p. 1, 2018.
- [27] M. Matta, I. Lunesu, and M. Marchesi, "Bitcoin spread prediction using social and web search media." in *UMAP workshops*, 2015, pp. 1–10.
- [28] S. R. Shah and A. Kaushik, "Sentiment analysis on indian indigenous languages: a review on multilingual opinion mining," *arXiv preprint arXiv:1911.12848*, 2019.
- [29] S. Kazhuparambil and A. Kaushik, "Cooking is all about people: Comment classification on cookery channels using bert and classification models (malayalam-english mix-code)," *arXiv preprint arXiv:2007.04249*, 2020.
- [30] G. Kaur, A. Kaushik, and S. Sharma, "Cooking is creating emotion: a study on hinglish sentiments of youtube cookery channels using semi-supervised approach," *Big Data and Cognitive Computing*, vol. 3, no. 3, p. 37, 2019.
- [31] S. R. Shah, A. Kaushik, S. Sharma, and J. Shah, "Opinion-mining on marglish and devanagari comments of youtube cookery channels using parametric and non-parametric learning models," *Big Data and Cognitive Computing*, vol. 4, no. 1, p. 3, 2020.
- [32] Alex, "Bitcoin tweets - 16m tweets," Nov 2019. [Online]. Available: <https://www.kaggle.com/alaix14/bitcoin-tweets-20160101-to-20190329>
- [33] H. C. Kraemer, "Kappa coefficient," *Wiley StatsRef: Statistics Reference Online*, pp. 1–4, 2014.
- [34] P. Joshi, "Elmo embeddings." [Online]. Available: <https://www.analyticsvidhya.com/blog/2019/03/learn-to-use-elmo-to-extract-features-from-text>
- [35] Y. Liu, C. Sun, L. Lin, and X. Wang, "Learning natural language inference using bidirectional lstm model and inner-attention," *arXiv preprint arXiv:1605.09090*, 2016.
- [36] H. Zhu, I. C. Paschalidis, and A. Tahmasebi, "Clinical concept extraction with contextual word embedding," *arXiv preprint arXiv:1810.10566*, 2018.
- [37] N. Ueda and K. Saito, "Parametric mixture models for multi-labeled text," in *Advances in neural information processing systems*, 2003, pp. 737–744.
- [38] K. Kowsari, K. Jafari Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text classification algorithms: A survey," *Information*, vol. 10, no. 4, p. 150, 2019.
- [39] S. Tan, "An effective refinement strategy for knn text classifier," *Expert Systems with Applications*, vol. 30, no. 2, pp. 290–298, 2006.
- [40] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE transactions on systems, man, and cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.
- [41] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, and D. Haussler, "Support vector machine classification and validation of cancer tissue samples using microarray expression data," *Bioinformatics*, vol. 16, no. 10, pp. 906–914, 2000.
- [42] I. Rish *et al.*, "An empirical study of the naive bayes classifier," in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, no. 22, 2001, pp. 41–46.
- [43] R. Kohavi *et al.*, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Ijcai*, vol. 14, no. 2. Montreal, Canada, 1995, pp. 1137–1145.
- [44] J. S. Farris, M. Kallersjo, A. G. Kluge, and C. Bult, "Constructing a significance test for incongruence," *Systematic biology*, vol. 44, no. 4, pp. 570–572, 1995.
- [45] R. M. Kozub, "An anova analysis of the relationships between business students' learning styles and effectiveness of web based instruction," *American Journal of Business Education*, vol. 3, no. 3, pp. 89–98, 2010.
- [46] H. Abdi and L. J. Williams, "Tukey's honestly significant difference (hsd) test," *Encyclopedia of research design*, vol. 3, pp. 583–585, 2010.
- [47] H. Park, N. Kim, and J. Lee, "Parametric models and non-parametric machine learning models for predicting option prices: Empirical comparison study over kospi 200 index options," *Expert Systems with Applications*, vol. 41, no. 11, pp. 5227–5237, 2014.
- [48] W. A. Jensen, J. B. Birch, and W. H. Woodall, "Monitoring correlation within linear profiles using mixed models," *Journal of Quality Technology*, vol. 40, no. 2, pp. 167–183, 2008.