The Impact of Tweet Sentiments on the Return of Cryptocurrencies: Rule-Based vs. Machine Learning Approaches

Peyman Alipour and Sina Esmaeilpour Charandabi

ABSTRACT

In an attempt to assess the appropriateness of the best-practice lexicon-based approaches as opposed to novel learning-based models to extract the sentiment of textual content in the context of the cryptocurrency market, the current study provides further insights into the association between digital activity and price movement of cryptocurrencies. Using a sample of Bitcoin and Ethereum trade data, this study compares the performance of Harvard IV-4 and BERT models in conjunction with the well-known machine learning classifiers. It examines to what extent learning-based sentiment models can enhance the price movement prediction, compared to lexicon-based approaches, and whether the prediction is improved or impaired by introducing different features as input to the classifiers. Results indicate that the contribution of the selected learning-based model varies across the two cryptocurrencies, and predictions are better in the absence of trade volume as an input feature to the classifiers.

Keywords: BERT, cryptocurrency, lexicon, machine learning.

1. Introduction

With growing interest in trading cryptocurrencies in financial transactions during the last decade, investors are becoming more interested in forecasting schemas, enabling them to predict the return on investment in the cryptocurrency market, which is subject to high volatility. Consequently, different predictive models and methods have been introduced since 2017, when the price of Bitcoin—the first decentralized cryptocurrency—experienced enormous hikes, and similar sudden hikes were observed across the price of all cryptocurrencies in the same timeframe (Charandabi & Kamyar, 2022). Such predictive approaches, which encompass a wide range of various methods, such as regression-based time series and machine learning (ML) models, are comprehensively discussed in recent surveys in the literature (Charandabi & Kamyar, 2021a, 2021b; Khedr et al., 2021).

The history of price is still considered among the main contributing factors to predicting the return on investment in the cryptocurrency market, as it is commonly used in predicting all financial markets. However, recent studies have also explored the explanatory power of some exogenous variables, such as the sentiment of the content provided through social media activity (Lamon et al., 2017). Given that social media activity reflects people’s most up-to-date perceptions of a given topic, sentiment analysis has turned out to be a valuable resource for constructing predictive models.

While lexicon-based approaches are still widely used for sentiment analyses (Alipour & Charandabi, 2023; Loughran & McDonald, 2011), Natural Language Processing (NLP) models are increasingly grabbing attention and gaining prominence for predicting price movements in financial markets (Aharon et al., 2022; Zou & Herremans, 2022). Unlike the rule-based nature of lexicon approaches, which makes them effective over only certain pre-identified keywords, the learning-based nature of NLP models grants more power to them in comprehending new entities in the context. The climax of enhancements in NLP models was probably in 2018 when a group of scientists from Google (Devlin et al., 2018) introduced the Bidirectional Encoder Representations from Transformers (BERT) model, which could learn more complex linguistic relationships.

Among existing social networking services, X formerly known as Twitter—has been a helpful data source for recent studies relating sentiment analysis to price volatility in the cryptocurrency market (Aharon et al., 2022; Zou & Herremans, 2022). In our previous study (Alipour &
we compared the performance of two different lexicon-based approaches in predicting the price volatility of cryptocurrencies through regression models. Due to the lack of sufficient content in the cryptocurrency literature in terms of implementing machine learning methods to predict price volatility and applying learning-based models in sentiment analysis, in this article, we will study a sample of tweets to gain further insights in two respects:

1. To what extent can learning-based sentiment models enhance the price movement prediction, compared to lexicon-based approaches?
2. Is prediction improved or impaired by introducing different features as input to the model?

2. Methodology

In accordance with our previous study (Alipour & Charandabi, 2023), the current analysis will also be conducted over trade data for Bitcoin (BTC) and Ethereum (ETH), which represent the top two widely traded cryptocurrencies in the market.

To answer the first research question, relevant tweets from the trade timeframe will undergo a sentiment analysis using two different approaches: Harvard IV-4 (HRV) as a commonly accepted lexicon-based model and BERT as a well-known learning-based model.

The outcome variable in this study will have a binary value associated with the negative or positive return in price. We will examine different ML-based classifiers to acceptably predict the return. Sentiment scores and prices from the previous day will always be present as input to models.

To answer the second research question, we will discuss the performance of the selected classifiers in the absence and presence of trade volume as a potential feature that is assumed to have a discriminatory power.

3. Data

3.1. Trade Data

Data for BTC and ETH trade, including historical prices and corresponding trade volumes, were collected from Coindesk.com during the three months toward the end of November 2021, when cryptocurrencies reached their highest price ever since their introduction to financial markets (Alipour & Charandabi, 2023). Figs. 1 and 2 depict fluctuations in BTC and ETH trade volumes in the sample, respectively.

3.2. Twitter Data

Following the explained guidelines in our previous study (Alipour & Charandabi, 2023), we used Python’s Sns scrape package to scrape associated tweets from X during the 90-day trade period, given “Bitcoin” and “Ethereum” as trade-related keywords. After extracting a sample of 45,000 tweets for each cryptocurrency, they underwent data pre-processing steps, including data cleaning (removal
of special signs, punctuations, and blank spaces), tokenization, and lemmatization.

HRV and BERT methods were applied to obtain sentiment scores using lemmatized tweets from the preprocessed data. Outputs of the HRV model indicate the degree to which a tweet has positive or negative sentiment, known as polarity score. Likewise, the BERT model returns the likelihood of positive or negative sentiment as the output. Given day as the level of the analysis in this study, the average of sentiment scores was aggregated from the extracted tweets for each corresponding day.

Figs. 1 and 2 display fluctuations in sentiment scores concerning HRV and BERT models and the return in price for BTC and ETH, respectively. A quick look at the trend of the variation of sentiment scores implies that they do not fluctuate according to price movements. Similarly, calculated scores through the HRV model follow a different direction than those of the BERT model.

As shown in Table I, on average, the BERT model assigns a higher probability to the positive sentiment when the return is positive, either for BTC or ETH. On the other hand, calculated scores by the HRV model do not differ significantly across the return direction. This observation implies the potential superiority of the BERT model, and it will be further examined through the classification experiment in the following section.

### 4. Model

#### 4.1. Measuring Return

The return is used as a binary outcome variable in the model. It has a “Positive” value when the closing price of a cryptocurrency on a given day is higher than that of the previous day, and it has a “Negative” value, and vice versa.

#### 4.2. Return of BTC

Based on our experiments through different ML classifiers over the BTC data, XGBoost appears to be the best-performing one, which isn’t vulnerable to overfitting. As shown in Table II, given AUC as an appropriate performance metric for this study, when trade volume is excluded from inputs to the model, the performance is improved, regardless of whether HRV or BERT is used to calculate the sentiment scores. However, the improvement is far larger (11.34%) when implementing BERT rather than using HRV (1.7%). It is also seen that BERT is the superior approach between the two sentiment models in either case; in the presence of trade volume among features, it enhances the performance to a very slight degree compared to HRV (0.44%), and when trade volume is absent, the enhancement is significantly at a larger level (9.90%). Effects on the classification performance through this feature engineering can be better visualized through the ROC metric in Figs. 3 and 4.

#### 4.3. Return of ETH

As per our experiments, SVM returned more robust ETH results than other ML classifiers. However, the model is subject to some degree of overfitting. As per AUC values in Table II, excluding the trade volume from inputs to the model still enhances the performance, but to a slighter degree than that of BTC. However, the improvement associated with BERT is still achieved at a larger ratio (8.90%) compared to HRV (5.76%). The noticeable difference, as also depicted in Figs. 5 and 6, is that unlike BTC, wherein
The Impact of Tweet Sentiments on the Return of Cryptocurrencies: Rule-Based vs. Machine Learning Approaches

Alipour and Charandabi

5. Conclusion

In an attempt to assess the appropriateness of the best-practice lexicon-based approaches as opposed to novel learning-based models to extract the sentiment of textual content in the context of the cryptocurrency market, the current study provides further insights into the association between digital activity and price movement of cryptocurrencies. First, we observed that there is no one-size-fits-all rule concerning selecting the best ML classifier to predict the return on investment. While XGBoost was realized as a well-performing model to predict the return on BTC trade through sentiment scores, SVM was a better selection for ETH, which has price volatility and trade volume at a narrower scale. Therefore, different predictive schemas are recommended to be examined for different cryptocurrencies to choose the right one.

Second, we provided empirical support that using BERT as a novel learning-based model to calculate sentiment scores, can result in a higher predictive power than HRV, a well-known lexicon-based model. However, we also noticed that the superiority of BERT depends on the type of cryptocurrency. As per the results, while BERT was the superior model for predicting the return on BTC data, HRV was preferred in ETH. This result is aligned with our previous study (Alipour & Charandabi, 2023), wherein HRV suggested enhanced results for predicting the volatility of ETH through the regression schema.

Despite this observation, we should also notice that fitting the SVM classifier over ETH data was subject to overfitting, which makes one cognizant of drawing definitive inferences about the performance of sentiment models.

Lastly, we realized that regardless of the type of cryptocurrency, type of the sentiment model, and type of the classifier, the inclusion of trade volume as an input feature impaired the performance of the prediction. This impact should be attributed to the noise imposed on the model in the presence of the trade volume, which affects the results of BTC to a greater extent, given the fact that BTC had the largest market cap and highest volatility in return, making its data more prone to bring noise into the model. It was also seen that regardless of the type of cryptocurrency and type of classifier, models supported by BERT improved to a greater extent when trade volume was excluded from the input features.

Given that ML classifiers typically perform well over large datasets, the sample size was the main challenge for us to acknowledge in this study. To capture the most visible volatility in price, we inevitably narrowed the sample down to 90 days, wherein the cryptocurrency market had more reflection in the media. However, extending the sample size to a broader timeframe confronts us with accessibility issues through non-commercial APIs to extract tweets from X, given the restrictions imposed on the platform to scrape data. Addressing this limitation would be a crucial step that researchers are encouraged to consider in future studies.

This study was intended to provide further insights by examining the functionality of the best practice ML classifiers, in conjunction with novel sentiment models, to predict the return in the cryptocurrency market (Charandabi, 2023; Charandabi & Ghanadiof, 2022; Zou & Herremans, 2022). With recent advancements in developing large language models (LLMs), another recommendation for future studies is to examine the accuracy of proposed models to predict the direction of return using sentiment scores suggested by LLMs.

Conflict of Interest

The authors declare that they do not have any conflict of interest.
REFERENCES


