PreBit: An NLP Enhanced Prediction Model for Bitcoin Price Using Twitter

Yanzhao Zou[1], Dorien Herremans[2]

[1] Singapore University of Technology and Design
yanzhao_zou@mymail.sutd.edu.sg
[2] Singapore University of Technology and Design
dorien.herremans@gmail.com

Bitcoin, with its ever-growing popularity, has also demonstrated unparalleled price volatility since its origin. This volatility, together with its decentralised nature, make Bitcoin highly subjective to speculative trading as compared to more traditional assets. We are interested in studying whether social media discussions from the general public on Bitcoin have predictive power for extreme future price movements. To obtain such discussion contents, a dataset of 5,000 daily Tweets (or the maximum number available that day) were collected from 2015 to 2021 containing the keyword "bitcoin". Previous studies have used the Bag of Words (BoW) approach with logistic regression to show that a link exists between Twitter sentiment and daily price on a short time frame of twenty-one days. In our study, sentence-level BERT embeddings pre-trained on financial lexicons were utilised in an attempt to capture not only the sentiment but also the contents of the tweets. By combining these embeddings with a Convolutional Neural Network, we closely examine the link between the public tweet contents and significant market movement over a much longer period of time. We also propose an ensemble of our NLP model and a baseline price model to explore how we can augment the performance of traditional price models with NLP based on Tweets for extreme price movement prediction.

Abstract to be presented in EURO2021,IFORS2021