Sentiment Analysis and Bitcoin Return Prediction Using FinBERT

Liangyu Chen
liangyu.chen@gatech.edu

Ruiqi Zhu
rich.zhu@scheller.gatech.edu

Mingxun Wang
mvw5171@gatech.edu

Zirong Yu
zroyu@gatech.edu

1 Abstract

In this study, we show that FinBERT model, compared to base-BERT model, performs worse on analyzing sentiments for non-financial text. By analyzing financial headlines, we show preliminary evidence that sentiment scores generated by FinBERT model have predictive power on future prices for traditional financial assets, such as stocks. Furthermore, we explore if FinBERT could be useful in analyzing sentiment scores based on social media posts and predicting nontraditional financial assets, such as Bitcoin. By conducting sentiment analyses on Bitcoin Tweets, we present evidence that FinBERT-generated sentiments provide predictive power on future Bitcoin price.

2 Introduction

BERT is a popular natural language processing (NLP) pre-trained model developed by Google, which is widely used for sentiment analysis. By further training the BERT language model in the finance domain, FinBERT is created. It is trained on a large financial corpus and fine-tuned for financial sentiment classification.

The goal of our study is to examine the efficiency and accuracy of the FinBERT model on both non-financial and financial text data. We also use FinBERT to conduct price predictions on stocks and Bitcoins. Our main research questions are as follows.

First, does FinBERT perform worse on non-financial data compared to BERT? Since FinBERT is a model specialized for analyzing financial text, whether it continues to performs well on non-financial data compaared to BERT remains unexamined. Thus, we evaluate the base BERT model and FinBERT model on non-financial data and compare their performance.

Second, does FinBERT perform better than base-BERT in analyzing sentiment of financial headlines? Even though FinBERT is expected to perform better, the improvement in performance is unknown. To explore this question, we use the FinBERT model to predict future stock price.

Finally, is FinBERT an effective tool in predicting future price of cryptocurrency such as bitcoin? We apply FinBERT on social media, more specifically Twitter tweets. We analyze retail investors' Twitter sentiment on Bitcoins and utilize sentiment information to predict future Bitcoin prices. Retail investor behaviors have attracted significant attention from market watchers and academics. Retail investors gain increasing power in shifting the market. Their participation during the COVID-19 pandemic led to dramatic market volatility and jaw-dropping price movements in so-called "MEME stocks", such as Bitcoin, Gamestop, and AMC. Compared to institutional investors, recent studies by Eaton et al. [1] have found that these investors are more likely to herd, harm liquidity when the market condition deteriorates, and drive higher volatility. Bradley et al. [2] also found that individual retail investors are more likely to share their trades on Reddit posts, which have been shown to have real-world impact on the stock market.

Retail investors play a huge role in the market and they communicate and share opinions through social media channels such as Twitter. Therefore, it is essential to incorporate social media information in predicting Bitcoin price. We hypothesize that sentiment derived from Bitcoin tweets will have predictive power to its future price for two main reasons. First, investors in the cryptocurrency market tend to be younger generation investors who are more likely to use social platforms to share their opinions. Second, individual investors might have more influence in cryptocurrency market compared to traditional equity market since institutional investors are not yet investing major stakes in cryptocurrency. Our results confirm that social media information is useful in predicting Bitcoin price.

3 Literature Review

Basant Agarwal and Namita Mittal [3] propose a sentiment analysis approach using machine learning methods in 2016, which serves as a useful guide for us to conduct sentiment analysis on financial data. In addition, Oscar Araque etc [4] develop a deep learning based sentiment classifier using word embedding models and a linear machine learning algorithm. He also introduce a new taxonomy for classifying the texts, which is useful in our analysis. Furthermore, Jeremy Howard and Sebastian Ruder [5] propose a Universal Language Model, which outperforms traditional sentiment analysis models such as BERT. That model could be used as a comparison with FinBERT model.

4 Dataset

4.1 Non-Financial Text: Disaster Tweets

We compare the performance of FinBERT and BERT on data that is outside the financial context. We use the disaster tweets dataset from the Kaagle website. The dataset is basically used to predict which Tweets are about real disasters and which ones are not.

The training dataset consists of 7,613 tweets and the testing dataset consists of 3,263 tweets in total. There are 55% non-disaster tweets and 45% disaster tweets in the training dataset. The distribution is well-balanced and the word count for non-disaster and disaster tweets share a similar distribution, making it a suitable dataset for our evaluation.

4.2 Financial Text: S&P 500 Price and Finance News Headlines

The dataset for FinBERT evaluation consists of two parts: S&P 500 constituents price information and finance news headlines in 2019.

We scrape the price information of S&P components from Wikipedia and Yahoo Finance, which contains 503 stocks in total and daily opening and closing price information in 2019. For the news headline data, we downloaded the financial news headlines of over 6,000 stocks from the Kaggle website, which contains the news headlines and release dates of major financial medias.

Using the two datasets, we construct a new data frame consisting of ticker number, news release date, opening price, closing price, headline, and 1-day return. In summary, we have 384 tickers and 41,235 headlines in the dataset, which is sufficient for our model performance comparison.

4.3 Bitcoin Price and Twitter

To obtain Bitcoin Twitter sentiment score, we collect 3,029,118 Tweets with related hashtags, such as bitcoin, btc, etc., during the period from February 2021 to April 2022. The data was downloaded from Google's Kaggle website, and contain important Tweet features, including post date, text, number of followers, number of favourites, number of friends, etc.

We have also collected Bitcoin price and S&P 500 ETF price for the same period as the Bitcoin Tweets data from Yahoo Finance. By combining the two datasets, we further analyze the prediction power of the Twitter Bitcoin sentiment on Bitcoin price.

5 Methodology

5.1 FinBERT on Non-Financial Text

In order to test the efficiency and accuracy of the FinBERT model on non-financial data, we use BERT-large-uncased as our baseline. This model uses 124-layer, 1024-hidden, 16-heads, 340M parameters BERT model with an uncased vocab and it is directly downloaded using the Transformers library. FinBERT model also comes from Prosus AI Transformers library.

We tokenize all the training test sentences and map the tokens to their word IDs. Then we combine the training inputs in a TensorDataset and create a 80-20 train-validation split. Training is performed on Google Colab and we use hyperparameters that are recommended in the BERT paper[4].

5.2 FinBERT on Financial Text

FinBERT is pretrained and available in Prosus AI. We download the model in Python and run classifications on our headlines. For each text headline, it has three softmax scores representing positive, negative and neutral sentiment. After deriving the scores, we assign a sentiment classification for each ticker based on the largest average softmax score of their new headlines on each day. In this way, for each ticker on each day, the FinBERT model would assign a classification result as "positive (+1)", "neutral (0)" or "negative (-1)".

Once we get the stock sentiments, we would like to know how the they correlate with stock returns. Therefore we compare the correlations of sentiment classifications with stock 1-day returns. Next, we split the whole dataset into two parts: training (90%) and testing (10%). After that, we build a Deep Neural Network (DNN) Model to fit the training dataset and forecast the closing price and evaluate the results. The DNN model contains 4 layers: 1 Normalization Layer, 2 Dense Layers with ReLU Activation, and 1 Dense Layer with Linear Activation.

5.3 Bitcoin Price and Twitter Sentiment

In order to generate sentiment scores in Twitter, we remove special characteristics, hashtags, emojis, website links, and other miscellaneous noises. Due to computational limitations, we randomly pick 1,000 tweets each day as the proxy of the sentiment score for the date. This “Bitcoin Tweets Sentiment Score” (sentiment score) is calculated as a sum of the "positive (+1)", "neutral (0)" or "negative (-1)". For instance, a score of 10 represents that on average 10% of the Bitcoin tweets in a certain day is positive after deducting negative tweets.

We calculate the daily return as follows:

$$Return_t = \frac{(AdjCloPr_t - AdjCloPr_{t-1})}{AdjCloPr_t}$$

where $AdjCloPr_t$ is the adjusted closing price for date t collected from Yahoo Finance.

In addition to sentiment scores generated by FinBERT, we also include $Return_{t-1}$ of Bitcoin to control for auto-correlation, $Return_{t-1}$ of S&P 500 ETF as a control for the benchmark and proxies of the degree of posts' influence in a day.

Feature	Description
Sentiment Scores	FinBERT-generated scores based on Bitcoin Tweets
Cryptocurrency Return on Day t	$Return_{t-1}$ of Bitcoin
Market Return on day t	$Return_{t-1}$ of S&P 500
Tweet Impact Score 1	Log of sum of user followers on day t-1
Tweet Impact Score 2	Log of sum of user friends on day t-1
Tweet Impact Score 3	Log of sum of user favourites on day t-1

6 Results

6.1 FinBERT on Non-Financial Text

The following table shows the training result on BERT and FinBERT model. Both BERT and FinBERT achieve a very good accuracy within 3 epochs and the validation loss keeps increasing after 3 epochs because of the over-fitting issues. We can easily see from the table that on average, BERT model gets a slightly higher accuracy than FinBERT model does on non-financial data.

	Epoch	Train loss	Val. loss	Acc.
BERT	1	0.56	0.47	0.81
	2	0.38	0.40	0.84
	3	0.31	0.42	0.85
FinBERT	1	0.57‘	0.43	0.80
	2	0.38	0.42	0.81
	3	0.33	0.44	0.81

6.2 FinBERT on Financial Text

For all datasets, we observe that 31.4% of records were labeled as "positive", 23.4% as "negative" and 45.2% as "neutral". Moreover, we find the overall correlation of the sentiment classifications and 1-day return was 0.276, showing that there was a weak positive correlation. The result matches our expectation.

Next, we apply the DNN model on the training sets and evaluate the results on the testing set. 48.3% of the predicted closing price are within 1% error, 76.6% are within 2% errors and 95.7% are within 5% errors. FinBERT performs very well and we use it for the prediction on Bitcoin prices.

The evaluation result of the FinBERT model and DNN predictions are presented below (Figure 1).

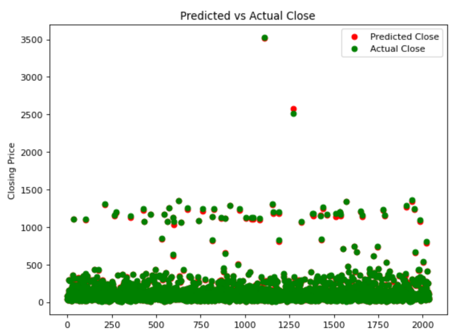

Figure 1: FinBERT Evaluation Result

6.3 Bitcoin Price and Twitter Sentiment

By analyzing Bitcoin tweets and price data, we find that sentiment has a positive correlation with Bitcoin return on both the same day and the next day. As expected, the correlation is more significant for the same day, which results in 0.257, compared to 0.0392 for the second day.

We also find that adding the sentiment score as a predictive feature could improve the prediction of Bitcoin's next-day return. The performance is measured by mean-squared-error. As shown in the following table, adding the FinBERT-generated sentiment scores could reduce MSE by about 5% in the Bitcoin return prediction model using random forest. However, when we conduct a similar analysis by linear regression, the effect seem to disappear. This might be due to the fact that the effect of sentiment in predicting Bitcoin return might be non-linear. Thus, machine learning models might have advantages in handling financial market prediction problems.

Model	Features	In-Sample	Out-of-Sample
Random Forest	w/o Sentiment	0.000203	0.00165
	w Sentiment	0.000201	0.00157
Linear Regression	w/o Sentiment	0.001184	0.00146
	w Sentiment	0.001181	0.00147

7 Conclusion

The goal of our study is two-fold. First, we compare the performance of FinBERT with base-BERT model on both financial and non-financial text data. The results show that FinBERT performs better on financial text data as expected but performs worse on non-financial text data in sentiment analysis task. Second, we use FinBERT in sentiment analysis to obtain sentiment score on tweets. Then we use the obtained score to predict future Bitcoin return and price movements. We find that adding the sentiment score as a feature could improve the prediction of Bitcoin's next-day return, but the performance improvement hinges on the machine learning model adopted.

Future work can more systematically compare FinBERT with base-BERT in other NLP tasks such as NER or machine learning translation. FinBERT can also be used to predict Bitcoin price in longer time horizons and be used in other cryptocurrencies such as Ethereum or Tether.

References

[1] Devlin, Jacob, et al., *Bert: Pre-training of deep bidirectional transformers for language understanding.* arXiv preprint arXiv:1810.04805 (2018).

[2] Gregory W. Eaton, T. Clifton Green, Brian S. Roseman, and Yanbin Wu. *Retail Trader Sophistication and Stock Market Quality: Evidence from Brokerage Outages.* Working paper, 2021.

[3] Basant Agarwal and Namita Mittal. *Machine Learning Approach for Sentiment Analysis.* Springer International Publishing, 2016

[4] Oscar Araque, Ignacio Corcuera-Platas, J. Fernando Sánchez-Rada, and Carlos A. *Enhancing deep learning sentiment analysis with ensemble techniques in social applications.* Expert Systems with Applications, 2017

[5] Jeremy Howard and Sebastian Ruder. *Universal Language Model Finetuning for Text Classification.* arXiv, 2018