Cryptocurrency price predictions using sentiment analysis – Milestone (section 4.)

VinUniversity

Course: Data Mining & Big Data Analytics

Instructor: Professor Khoa Doan

Vu Nguyen (V202000145)

1. Introduction
    a. Problem statement

Nowadays, with a fast-paced world of cryptocurrencies, market dynamics are heavily influenced by a mix between traditional financial indicators and the powerful swings of public sentiment. The inherently unstable nature of cryptocurrency prices means they react strongly to the collective emotions and viewpoints shared across different media channels. This complex relationship between market performance and public conversation offers a unique set of challenges and opportunities for investors and analysts dedicated to refining their price movement predictions.

Hence, predicting price trends specifically related to cryptocurrency is an extremely difficult task because of various factors. Moreover, typical classification models for financial analysis usually fail to capture all of the influences, especially those which are rooted in public sentiment. This shortfall can lead to an occlusion in capturing the whole context in the price analysis. Therefore, given the critical role that investors sentiment plays in influencing market dynamics, there is a persuasive need for models that incorporate with the sentiment. This approach promises to improve predictive accuracy among the existing methods. As a consequence, this project aims to create a sophisticated predictive model that uses public sentiment to forecast the price movement more precisely.

    b. Why use sentiment analysis?

The main factor that drives the price of cryptocurrencies is obviously the investors. Thus, using sentiment analysis serves as a powerful tool to capture public opinion and emotional responses, which plays a pivotal role in shaping the domain. The following underscore the advantages of integrating sentiment analysis to the price predictive model:

- Immediate response to news.
- A feature to enhance the classification model.

By addressing these aspects, this project aims to build a more robust model for cryptocurrency price prediction, enhancing accuracy through the integration of sentiment analysis. This approach not only aligns with current technological advancements in data analytics but also caters to the unique characteristics of the cryptocurrency market.

2. Proposed method
    a. Existing method

Actually, there has been excessive research in regard to predictive price model using sentiment analysis. Thus, this section will focus solely on the aspects that I would like to enhance within the current methodologies.

- The conference paper "Cryptocurrency price prediction using Twitter sentiment analysis" (Haritha & Sahana, 2023) suggests using GRU model to predict the price because of its ability to handle the vanishing gradient problem from using RNNs methods. Also, the authors use a fine-tuned BERT model to predict the sentiment because they believed that using transformer encoder such as BERT would be best to tackle the NLP problems, specifically in the financial domain. Additionally, VADER was used as explained in the text due to its belief that it is outperforming human raters. The evaluation metric used in the paper was MAPE (Mean absolute percentage error) which illustrates the difference between the predicted and actual data. The paper achieved an average of 9.45% MAPE for sentiment prediction using finBERT and 3.6% MAPE for the price prediction (Bitcoin price).

Table 3. Comparative results of Sentiment Prediction Methods

Time Period	**Mean Absolute Percentage Error**					
	Bi-LSTM		**GRU**		**FinBERT-NN**	
	Real-time	***Test data***	***Real time***	***Test data***	***Real time***	***Test data***
5 August 2022 - 5 September 2022	12.32	11.34	11.47	9.32	9.91	8.93
5 September 2022- 5 October 2022	12.01	11.05	11.57	9.29	9.78	9.19
Average	12.17	11.20	11.52	9.31	9.85	9.06

Table 4. GRU Price Prediction Results

Time Period	**Mean Absolute Percentage Error**	
	Test data	***Average***
5 August 2022 - 5 September 2022	3.44%	3.6%
5 September 2022- 5 October 2022	3.77%	

Figure 1: Result comparison between Bi-LSTM, GRU for price prediction and result of FinBeRT.

- In "A deep learning-based cryptocurrency price prediction scheme for financial institutions" (Patel, Tanwar, Gupta, & Kumar, 2020) suggested that using a combination of LSTM and GRU model returned a best result. Notice that this paper only focuses on price prediction but not using the sentiment analysis as a feature in their model.

Table 5
Results of LSTM and Proposed Approach for 7-days prediction window

Model	Currency	MSE	RMSE	MAE	MAPE
LSTM	Litecoin	286.1674	16.9164	14.7071	15.9441
	Monero	523.7109	22.8847	19.6743	19.1254
Proposed	Litecoin	20.7219	4.5521	3.8135	4.9407
	Monero	409.8096	20.2437	19.5513	19.3493

Figure 2: LSTM vs LSTM + GRU for price prediction

b. Proposed method & experiment

From the previous mentioned papers, I'd like to test out some of the following methods:

- Objective: to re-implement and to test the combination of LSTM + GRU model as suggested by Patel, Tanwar, Gupta, & Kumar for Bitcoin price prediction.
- Comparison base: Evaluate the performance of the GRU model that was used by Haritha & Sahana in their conference paper. (this because they achieved a better MAPE value in predicting the Bitcoin's price)
- Methodology:
  - Incorporate sentiment analysis from cryptocurrency community on investing.com forum.
  - Integrate sentiment data to fine-tune the LSTM + GRU model.
  - Maybe run some regularization (L2) to prevent overfitting and ensemble to improve accuracy if it is necessary.
- Performance comparison:
  - Compare the results of LSTM + GRU against GRU with and without the sentiment analysis integration.
  - Run some trials to see if the model is enhanced or not when using sentiment analysis.
- Baseline: The baseline would be the re-implementation of the methods from 2 papers then I will go from there to improve it if it is possible.

c. Dataset

Since I do not have access to the data that the authors from above papers use, as a consequence, I will use the data extracted from the site “investing.com” to get Bitcoin’s price (open, close, high, low, price, volume) on a daily basis (7 days' timeframe). For sentiment prediction, I will use the data which is provided by Kaggle (*Bitcoin_tweets.csv*), label it with VADER to pre-train the BERT model and fine-tune it using the real-time comments investing.com’s forum. The reason that motivated me to use investing.com is because the comments from their community do not have too much of “meme” so that it can reduce the neutral sentiment comparing with other sources like reddit or twitter.

3. Evaluation metrics

To give a “fair and square” comparison with the above papers and my experiments, I will use MAE, MSE, MAPE and RMSE to evaluate the proposed scheme.

- $MSE = \frac{1}{N}\sum_{i=1}^{N}(Y_i - Y_i')^2$
- $MAE = \frac{1}{N}\sum_{i=1}^{N}|Y_i - Y_i'|$
- $MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{Y_i - Y_i'}{Y_i}\right|$
- $RMSE = \sqrt{MSE}$

Where *Yi* is the actual data, *Y’i* is the predicted data and n is the number of observations.

4. Progress Milestone

From the last proposal, I have been working on building the LSTM + GRU model and finished with crawling data from the site “investing.com” using BeautifulSoup. The features I aim to crawl are *'open', 'close', 'high', 'low', 'volumeUSD', 'average', 'positive', 'negative', 'score'*.

	cryptoId	start	duration	open	close	high	low	volumeUSD	source	count	average	std	positive	negative	score
0	btc	1716508800000	86400000	67828	68535	69185	66714	1.186800e+11	reddit	2500	2.75	0.65	0.55	0.20	0.35
1	btc	1716422400000	86400000	69297	67924	70036	66671	1.113400e+11	reddit	2600	2.80	0.66	0.56	0.21	0.34
2	btc	1716336000000	86400000	70145	69225	70600	69049	1.161000e+11	reddit	2700	2.85	0.67	0.57	0.22	0.33
3	btc	1716249600000	86400000	71370	70104	71739	69266	1.403700e+11	reddit	2800	2.90	0.68	0.58	0.23	0.32
4	btc	1716163200000	86400000	66222	71418	71457	66140	9.769000e+10	reddit	2900	2.95	0.69	0.59	0.24	0.31

Figure 3: Lastest data crawling from investing.com

I have not tested with the crawled dataset yet but only with the 2019 dataset. However, with 2019 dataset, I achieved MAE ~ 0.05, MSE ~ 4.98, RMSE ~ 2.23, MAPE 2.24 which indicates that the model seems to have a decent performance but there are some large errors in the predictions.

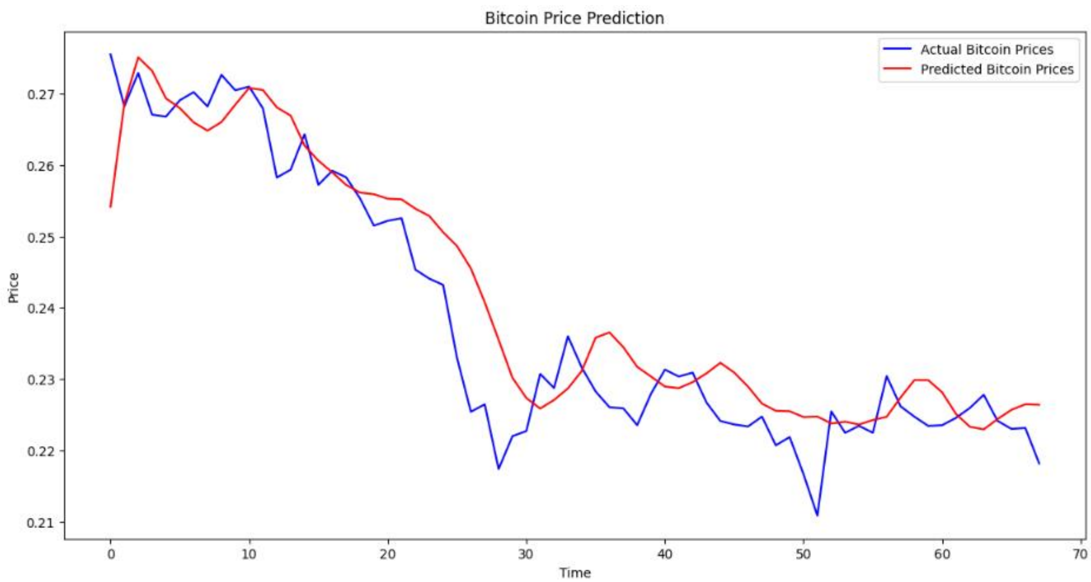

Figure 4: LSTM + GRU on 2019 dataset.

For the next steps, I will try tuning the model with different hyperparameters such as experimenting with number of layers/units, try different activation functions and most essentially is to crawl more data, probably in a smaller time scale (such as by every 15 minutes) because they are believed to be more difficult to predict.

5. Conclusion

In the fast-evolving realm of cryptocurrency markets, the convergence of traditional financial indicators with the profound sway of public sentiment underscores the intricate dynamics influencing price movements. The integration of sentiment analysis emerges as a pivotal strategy to navigate this complexity, offering a pathway to refine predictive models and enhance their accuracy. By incorporating sentiment data, these models gain a nuanced understanding of market sentiment, enabling investors and analysts to glean valuable insights and anticipate price trends more effectively. Leveraging methodologies from prior research, such as the combination of LSTM and GRU models, alongside innovative approaches like sentiment analysis from platforms like investing.com forums, the proposed experiment seeks to evaluate the efficacy of different model configurations in forecasting Bitcoin prices. Through rigorous evaluation using metrics like MAE, MSE, MAPE, and RMSE, this research endeavors to contribute to the development of robust predictive models that can adapt to the dynamic landscape of cryptocurrency trading, empowering stakeholders with actionable insights for informed decision-making in this rapidly evolving domain.

6. Reference

- Patel, M. M., Tanwar, S., Gupta, R., & Kumar, N. (2020). A deep learning-based cryptocurrency price prediction scheme for financial institutions. Journal of Information Security and Applications, 55(102583). https://doi.org/10.1016/j.jisa.2020.102583
- Haritha, G. B., & Sahana, N. B. (2023). Cryptocurrency price prediction using Twitter sentiment analysis. In D. C. Wyld et al. (Eds.), Proceedings of the NIAI, MoWiN, AIAP, SIGML, CNSA, ICCIoT - 2023 (pp. 13-22). CS & IT - CSCP 2023. https://doi.org/10.5121/csit.2023.130302
- https://www.kaggle.com/code/codeblogger/bitcoin-sentiment-analysis/