AIHACK Crypto

Shee Toh Benedict LEE, Wei Loon CHENG, Aiyun GUO

February 2021

Abstract

Given the constantly growing interest in cryptocurrencies such as Bitcoin, there is much value in developing a model for algorithmic trading of these volatile cryptocurrencies. In view of time constraints, our team developed a predictive model for close BTC-USD prices to serve as a starting point for those who want to develop a full-fledged trading strategy. Besides time-series data, sentiment scores of Tweets relating to Bitcoin and the relative popularity of Bitcoin searches on Google Trends were included independently as features in our multi-dimensional Long Short Term Memory (LSTM) model. We found that there was close agreement between our predicted prices and the actual close prices for all 3 models.

1 Motivation and Objective

The rise (and fall) and subsequent rise of Bitcoin has captivated aspiring traders across the world. The very fact that Bitcoin has seen such widespread mainstream exposure has made cryptocurrencies a hot topic of discussion. Recently, Elon Musk acquired $1.5 bn worth of Bitcoin, and confirmed that it would be accepted as payment in the future. There is thus much value in developing a model for algorithmic trading of these volatile cryptocurrencies.

Our team aims to develop a predictive model for Bitcoin prices, in the hopes that it will serve as a starting point for other traders who want to develop a full-fledged trading strategy.

2 Previous Work

A quick Google search reveals that other data enthusiasts have also tried to develop a predictive model for Bitcoin prices. However, most models take as input daily close prices to predict close prices for the following day.

Our team set out to develop a multi-dimensional Long-Short Term Memory (LSTM) model, which takes into account both open and close prices to predict close prices for the following day. This model is more robust and would lead to more accurate predictions.

3 Data Engineering and Processing

In addition to the spot trading dataset provided by Kaiko, we scraped hourly and daily BTC-USD price data from Yahoo! Finance using the APIs available. [1]

We also wanted to investigate whether Bitcoin sentiment (on Twitter) can better predict the close prices. Thus, we cleaned up data obtained from a Kaggle dataset [2]. The Kaggle user scraped all tweets mentioning 'bitcoin' from 1st August 2017 through 21st January 2019. The tweets with negative and positive sentiment were then assigned a score from -1 to 0 and from 0 to 1 respectively. The average of the negative sentiment scores was calculated for each hour, as was the average of the positive sentiment scores. We added up these averages to obtain an 'overall' sentiment for 'bitcoin' for each hour and each day.

To further complement our project, we looked into the relative popularity of 'bitcoin' searches daily on Google Trends for the year 2020.

4 Methodology

Initially, we experimented with a single dimension LSTM model which uses the current close price to predict the close price for the following day. The model had one input layer, 1 LSTM neural layer, and one output layer.

[1] https://sg.finance.yahoo.com/quote/BTC-USD?p=BTC-USD

[2] https://www.kaggle.com/jaimebadiola/bitcoin-tweets-and-price]

To predict the closing price for the following day using multiple inputs, we used a multi-dimensional LSTM model. We first trained the model on open prices, close prices and adjusted close prices. Subsequently, we trained the model on open prices, close prices, adjusted close prices, and Twitter sentiment scores from the Kaggle dataset. Lastly, we trained the model on open prices, close prices, adjusted close prices, and the daily relative popularity of 'bitcoin' searches on Google Trends [3].

5 Results and Discussion

Feature inputs always include open price, close price, and adjusted close price unless otherwise indicated. Other feature inputs such as Twitter sentiment scores and relative popularity of 'Bitcoin' searches on Google Trends were included in subsequent analysis.

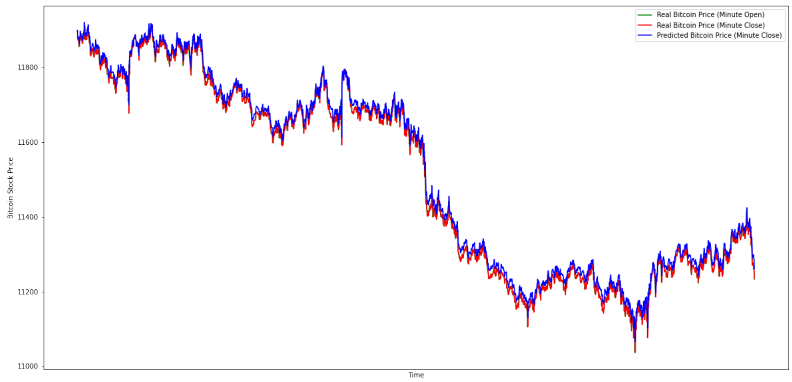

Figure 1: Predicted vs Actual Bitcoin Trend (Kaiko Cryptocurrency Dataset)

As a starting point, we used close prices of previous days to predict the close price of the following day. However, we felt that this could be improved on using multiple feature inputs. Hence, we decided to include open prices and adjusted close prices as parameters in our model. Figure 1 shows that there is a very good agreement between predicted prices and actual close prices.

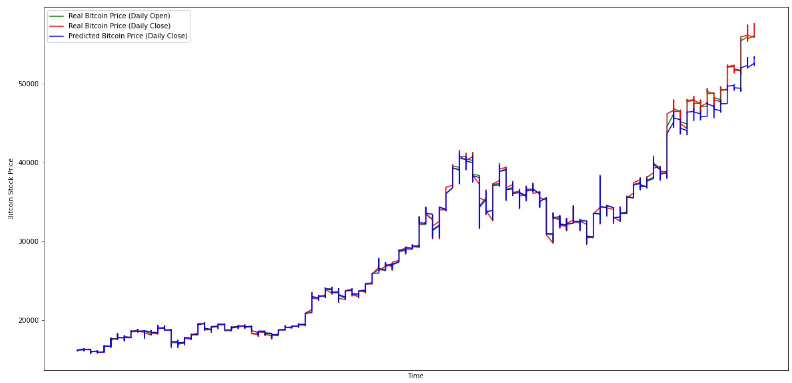

Figure 2: Predicted vs Actual Bitcoin Trend (Yahoo Finance 2019 - 2021)

In an attempt to improve our model, we scraped the last two years of BTC-USD data from Yahoo! Finance. We believed that our model would be able to make more accurate predictions given a larger dataset. However, the price activity of Bitcoin was relatively low in the first half of the timeframe. This caused the model to be less sensitive to large fluctuations. As seen in Figure 2, in the recent few months where prices shot up, we observed that our model did follow the rising trend, but with a smaller magnitude.

[3]https://trends.google.com/trends/explore?date=2020-01-01%202020-12-31&q=bitcoin

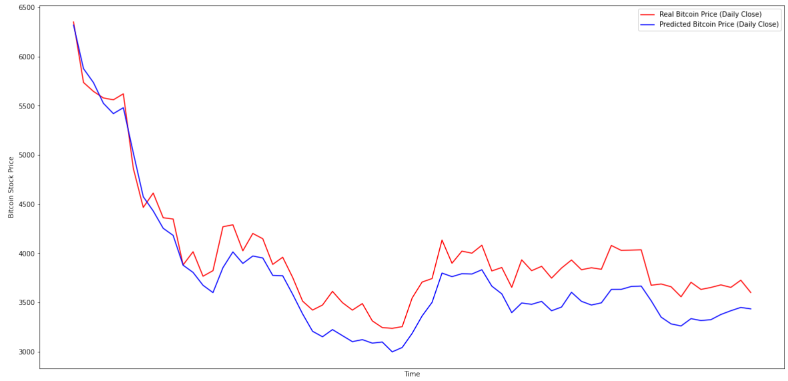

Figure 3: Predicted vs Actual Bitcoin Trend (With Sentiment Scores)

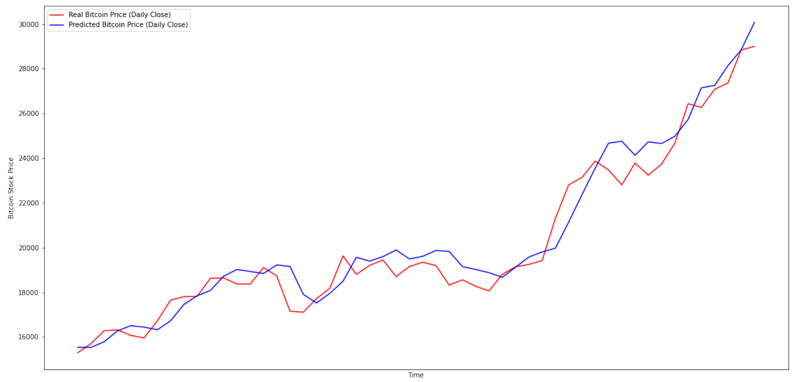

Figure 4: Predicted vs Actual Bitcoin Trend (With Google Trends data)

The accuracy percentages are shown in Table 1.

Datasets	Kaiko	Yahoo Finance	Twitter	Google Trends
Accuracy	99.96%	99.00%	93.68%	96.98%

Table 1: Accuracy Measure for each Dataset

Given the sharp bursts of interest in Bitcoin and associated volatility, we suspected that there was a strong correlation between public sentiment and Bitcoin price trends. We thus hypothesised that sentiment analysis related to bitcoin using Tweets will help in accuracy in predicting next day close price. Figure 3 shows a very good fit for the close prices generated by our model when sentiment data from Tweets related to Bitcoin was included. Similarly, we included the relative popularity of Bitcoin searches on Google Trends in another instance of our model. This also produced good results as seen in Figure 4.

6 Conclusion

In summary, we developed a multi-dimensional LSTM model to predict close prices for BTC-USD. The initial model trained on the data from Kaiko was promising, and was further improved with additional data from Yahoo! Finance. To improve the model, more feature inputs such as open price and adjusted close price were included. As Bitcoin prices are likely correlated with public sentiment, we also trained our model on Twitter sentiment scores and relative popularity of Bitcoin searches on Google Trends.

In the future, we aim to explore other models to analyse the dataset and predict future prices. We hope to further fine-tune the model based on other feature inputs and parameters that are relevant to Bitcoin. Currently, our model uses one previous timestep to predict the next. In our extension, we hope to predict the next timestep using the past N steps, where N can be fine-tuned as well.

A Appendix

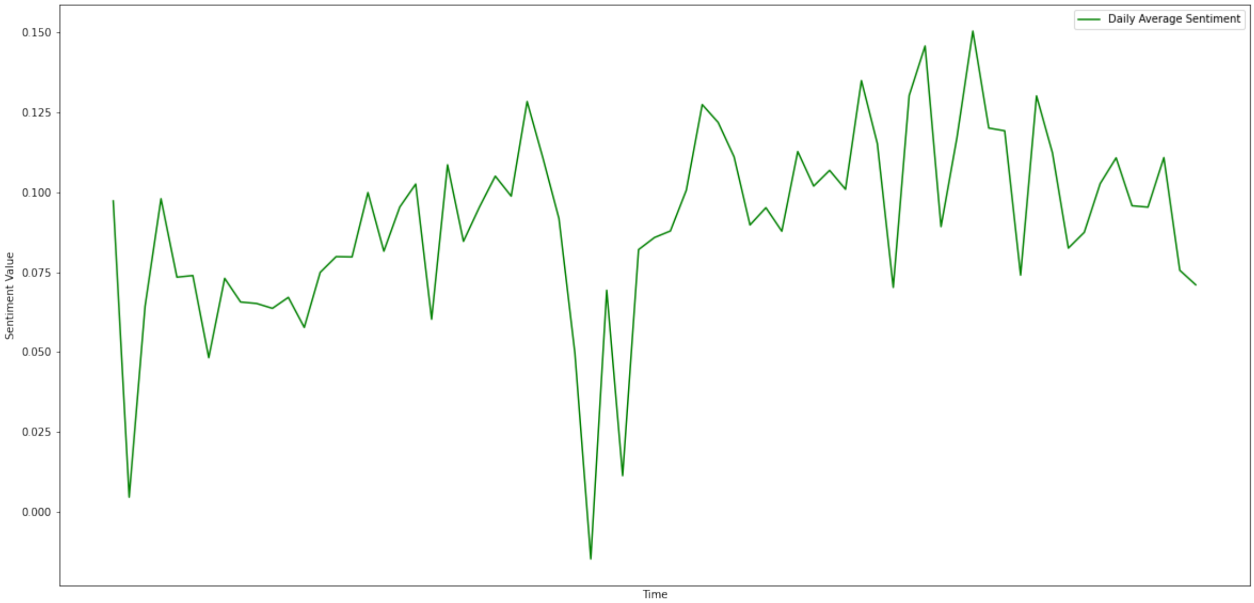

Figure 5: Bitcoin Sentiment Scores

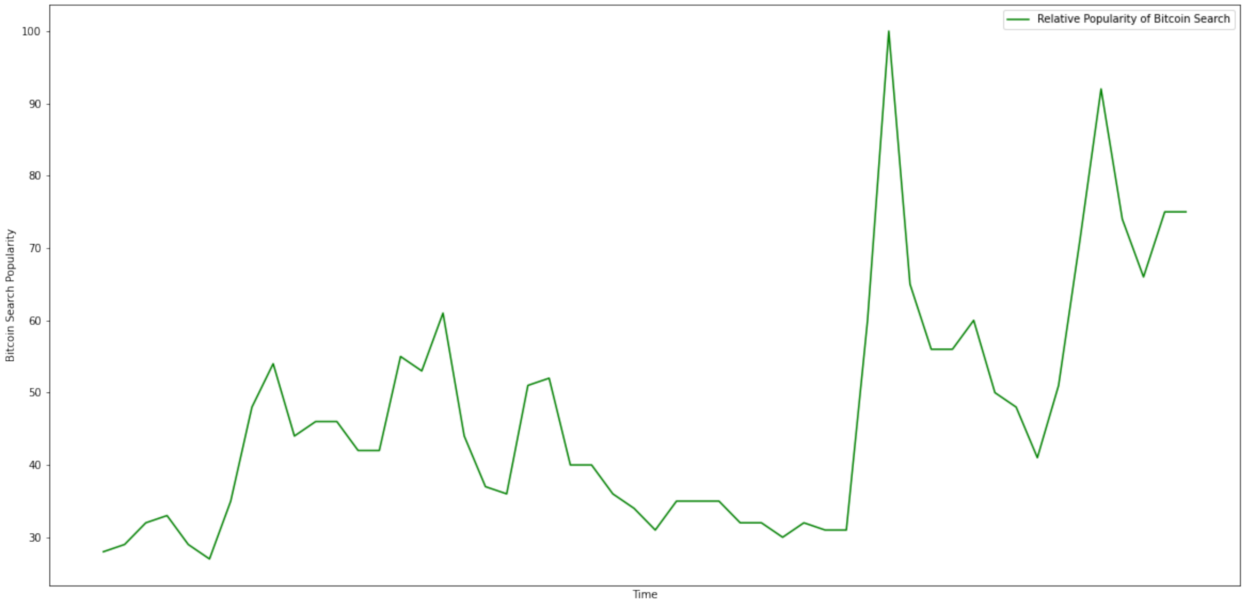

Figure 6: Relative Popularity of Bitcoin Searches on Google Trends