ANAL 590 Final Project

Bitcoin Price Prediction

Shiying Sun, Ruizhe Li, Ning Hu, Dongru Jia

Bitcoin Price Prediction Project Report

Abstract

Our project used web-scraping to obtain Bitcoin historical data, NASDAQ index historical data, volatility data. Then obtained feature engineering by adding 12 technical indicators. Then we explored different kinds of methods to do the price prediction including traditional trading strategies, time series machine learning models and LSTM deep learning models.

Introduction

Compared with traditional financial instruments, cryptocurrency is very difficult to predict due to the lack of indicator data. On the other hand, the fact that cryptocurrencies do not heavily depend on news, the market, crude oil, metals and other factors makes cryptocurrencies more suitable for prediction using deep learning than stocks.

Related Work

After reading tons of papers, we think LSTM is the most suitable model to do this prediction. Most related projects are done by using reinforcement learning to build an environment to mimic the trading process which is not covered in this course.

Datasets

Our primary dataset is obtained by using CoinBase API to scrape historical data containing date, open price, close price, highest and lowest price, market cap and volume. Other supporting stock market datasets are obtained from Yahoo Finance. The datasets are merged on date. There are empty rows in stock market datasets due to the fact that the stock market does not operate on weekends and holidays whereas the Bitcoin market never closes. The NA values are filled by the mean of the previous open day and the open day afterwards.

Methods

- VWAP:

We used one of the most important indicators in trading, VWAP, to mimic the trend of Bitcoin price. The volume weighted average price (VWAP) is a trading benchmark that gives the average price a security has traded at throughout the day, based on both volume and price.

- Time Series:

We used time series models on the Bitcoin prices as baseline models. We implemented both ARIMA and Prophet time series models and the resulting graphs are shown in the Results section. Prophet is an automatic time series forecasting model developed by Facebook.

- GAN:

The GAN architecture used a two-layer dense network (GRU and CuDNNLSTM) and one dense output layer with relu activation.

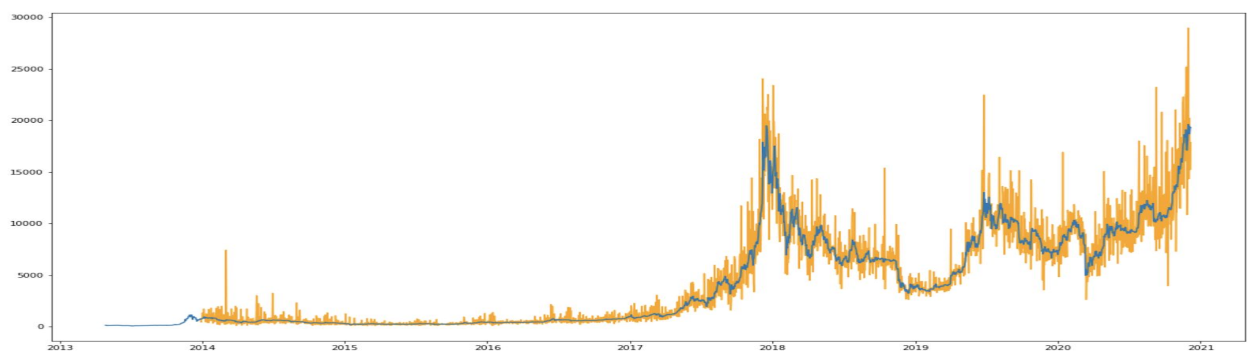
- LSTM:

We built an LSTM recurrent neural network with a window method to predict the last value of a sequence of values. We prepared the data into windows with a length of sixty days. The model takes in the previous sixty days of Bitcoin price to predict the price of the next day, which means there will be input units in the input layer. This method only predicts the value of the next day. So, in order to predict the future price, we applied the Seq2seq method to this model which uses predicted value as input in the next step.

Results

- VWAP:

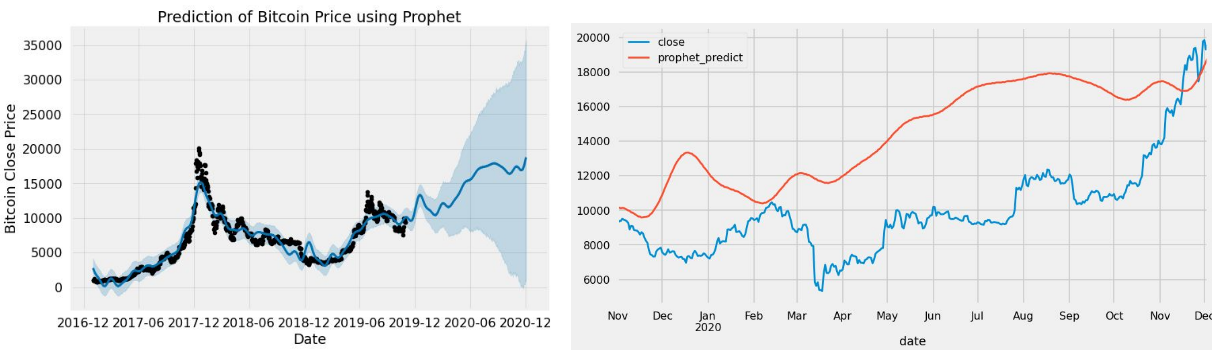
Due to the extreme instability of the volume of the bitcoin market, the VWAP indicator shows a strongly volatile trend of bitcoin price.



- Time Series:

The basic time series models: ARIMA and Prophet, did not fit well with test data and the root mean square errors and mean absolute errors were all really large. Although the Prophet model had higher errors than the ARIMA model, it successfully predicted the increasing trend in testing data and also predicted the highest price in December. Due to the lack of seasonality and trend in Bitcoin prices, we realized that predicting the prices simply based on the prices was not ideal. We then tried to use other models and implement more indicators.

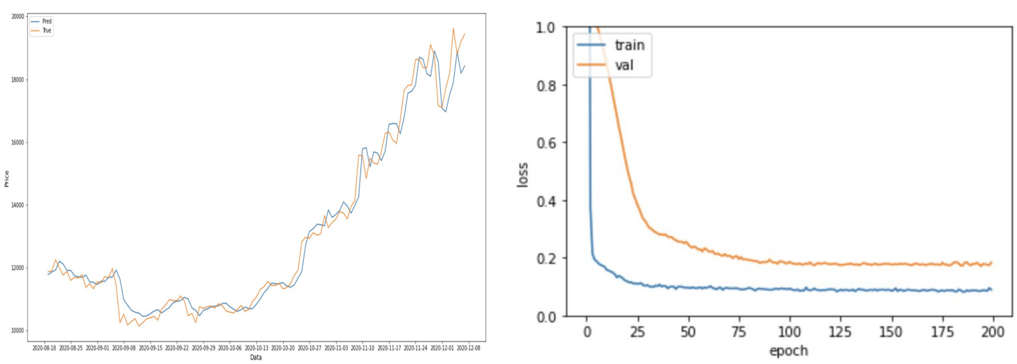
ARIMA Model, RMSE: 3175, MAE: 2132



Prophet Model, RMSE: 5078, MAE: 4569

- GRU:

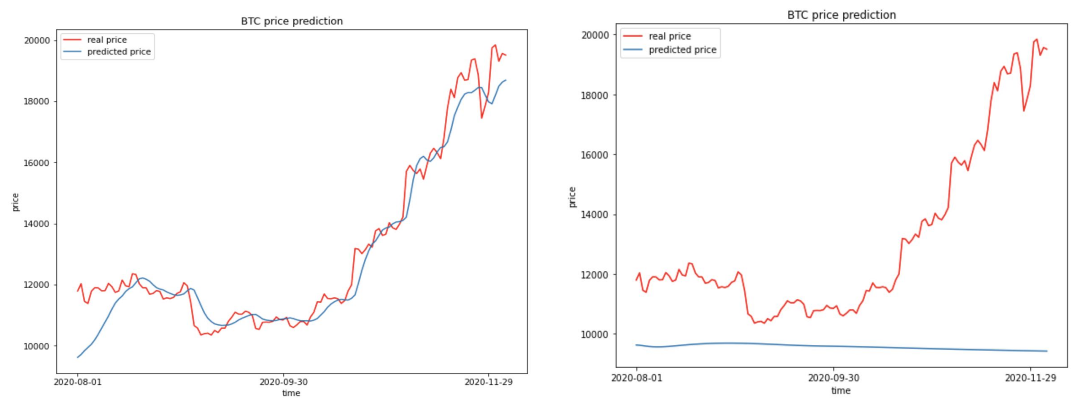
The loss of GRU cannot go lower than 0.18 on the validation set with look back equals 1.



GRU Model, RMSE: 406, MAE: 270

- LSTM:

The result of this approach looks pretty similar to the actual price since neural networks can fit even the most wiggly lines really well after enough epochs. In the seq2seq method, we take the output of the current step as the input for the next input. However, the result is very inaccurate. The predicted price is always around the average price of the training set. The downside of this method is that it can't predict the price sixty days but only one day ahead.



Lagged 1 (left) Seq2seq (right)

Discussion of Results

By trying out multiple approaches, we noticed that predicting the future with solely historical price is fairly unreasonable. Because essentially tomorrow's price will never be completely or even predominantly determined by today's price. In other words, there is no guaranteed pattern that is strong enough to outweigh the influence of other factors, such as a country's crypto currency policy or the release of a better GPU, which can substantially speed up the process of Bitcoin mining.

Conclusions

Nonetheless, the models we constructed still captured some juice in forecasting. To be specific, the GRU and LSTM model which uses the training data within a sixty days window managed to obtain persistence that is a term in price forecasting when a model uses today's price to predict that of tomorrow. And it is often served as the baseline predicting strategy in any forecasting task. In the ARIMA and LSTM model which was adapted into seq2seq, the model once again managed to make the best prediction on future price, which results in the lowest loss, by guessing the average of maximum and minimum price that occurred in the training set. To move forward in price forecasting with Deep Neural Network, one should not solely use price in the past, but consider adding more informative features, which both gives the model more interpretability and takes the precious advantage of Deep Neural Network's capability of extracting useful feature interactions.

Works Cited (APA Format)

K. Zhang, G. Zhong, J. Dong, S. Wang, and Y. Wang, "Stock Market Prediction Based on Generative Adversarial Network," Procedia Computer Science, vol. 147, pp. 400–406, 2019.

S. Siami-Namini and A. S. Namin, "Forecasting Economics and Financial Time Series: ARIMA vs. LSTM," pp. 1–19, 2018.