Bitcoin Price Time Series Prediction Using ARMA & LSTM Model

Hangran Zhang: A53256363
Yichen Zhang: A53288290

Abstract
The bitcoin price is now dramatically decreasing since last year December which make the prediction of the bitcoin price at the end of this year and in several years becomes a hot topic nationwide . The paper will conclude how we use ARMA and LSTM model separately to predict the bitcoin price in short and long term, and compare the performance of each model by MSE (mean squared error). The project will include DATASET ANALYSIS, ARMA MODEL, and LSTM PREDICTION.

Introduction
The value of bitcoin has seen significant losses over the last week, dropping to its lowest price since October 2017. A hack on a major South Korean exchange, as well as a new study suggesting it's 2017 highs were artificially inflated, saw the world's most valuable cryptocurrency fall below $6,000 to an eight-month low. The volatile cryptocurrency's price has shifted wildly ever since mid-December - when it hit a record high of more than $19,850 (£14,214) - with frequent heavy drops and speedy recoveries. [2] As a cryptocurrency, a form of electronic cash, bitcoin has become a hot spot in recent years.It is a decentralized digital currency without a central bank or single administrator that can be sent from user-to-user on the peer-to-peer bitcoin network without the need for intermediaries.[1] In this project, we first obtain the price data from Bitfinex, and then we design several functions to automatically get bitcoin price with one minute interval of a whole day. After that, we use ARMA and LSTM model to analyze the data and predict respectively. Finally, we analyze the result from the two model and propose some insight into the result.

Data Analysis
A. Obtain the price data from Bitfinex.
We use the Bitfinex website to obtain the dataset of the bitcoin prices. Bitfinex provides an overview of the state of the market. It shows the open price, close price, highest price and lowest price within the trade time frame. It also includes information such as daily volume and how much the price has moved over the last day.In order to get the bitcoin trade prices of every minute during a long time frame, We use pycurl toolbox in python to deal with the dataset and make the analysis on it. We first define a getpricefunction, which takes the millisecond time-stamp and the coin symbol as inputs, and returns the trade prices and volume of the following 720 minutes from Bitfinex. Then we define a oneday function to change the start time we want from numerical date into the standard time-stamp format, and call the getprice function two times to get a whole day's trading information per minute and write the final dataset to a csv file. From what we have done above, we can get a whole day's bitcoin price dataset simply by calling the oneday function, and we use panda toolbox to get the dataset in python notebook in order to analyze it easily. The head of the dataset looks as follows.

MTS	OPEN	CLOSE	HIGH	LOW	VOLUME
2018-12-14 00:00:00	3357.700000	3359.5	3360.0	3357.7	15.391929
2018-12-14 00:01:00	3359.600000	3360.0	3360.0	3359.6	3.663016
2018-12-14 00:02:00	3359.934622	3360.0	3360.0	3359.9	10.773534
2018-12-14 00:03:00	3360.000000	3360.9	3360.9	3360.0	11.601627
2018-12-14 00:04:00	3361.000000	3361.0	3361.0	3360.9	5.011997

Figure1: dataset of bitcoin price per minute

B. Analysis of the price data

We set the column of date as the index, in order to obtain the price of every minute for the following work conveniently. Many of our modeling processes are based on the large number theorem and the central limit theorem preconditions that we have learned from class. The large number theorem and the central theorem all require samples with the same distribution, which is equivalent to the stability of the time series. If it is not satisfied, many of the conclusions obtained are unreliable. Therefore, we define some function to analyze the stationary stability of the data series. First, we define a draw trend function to calculate the rolling mean averages and weighted mean averages of in order to smooth out the data to see the base trend of it. In order to quantify the correlation between the point at t and point at t + k (k is the lag) with respect to expectation, autocorrelation is used. And we also use partial autocorrelation to describe the direct effect of a value at t on value at t + k, ignoring the values between them. So we define a draw acf pacf function to get the autocorrelation and partial correlation.From the figure, we find out that both the auto-correlation and partial correlation coefficients have trailing characteristics, and they all have obvious first-order correlations, so we set p = 1 and q = 2. Below we can use the ARMA model for data fitting.

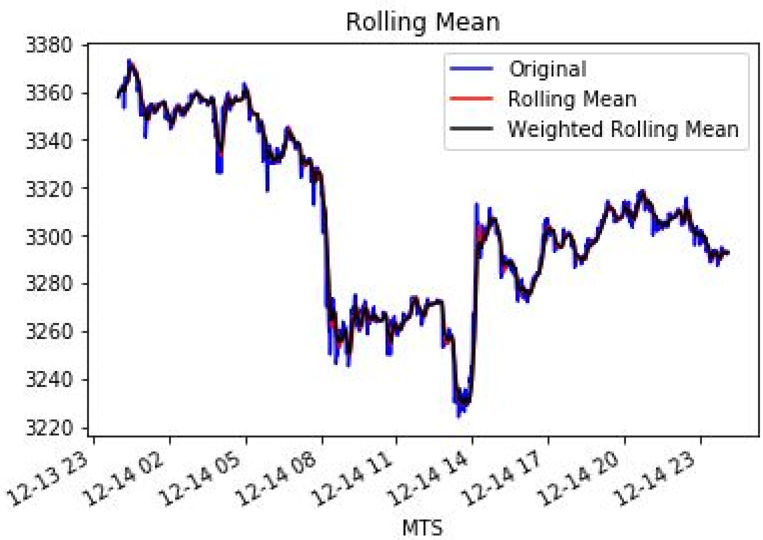

Figure 2: rolling mean of dataset

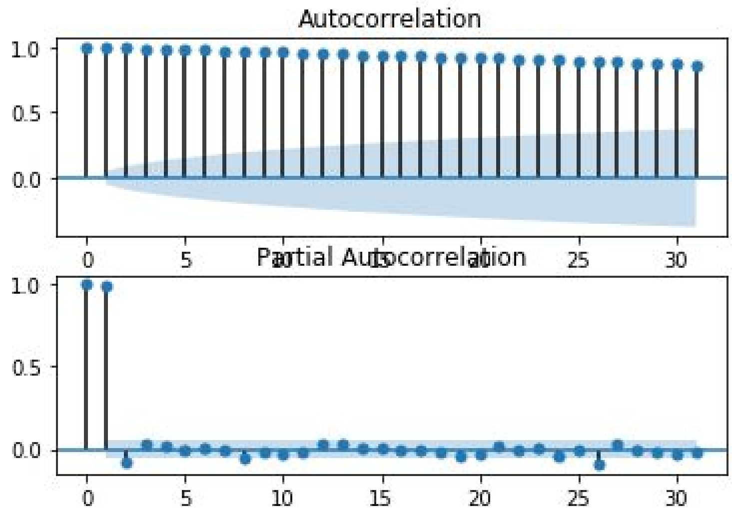

Figure 3:. autocorrelation and partial correlation of dataset

ARMA MODEL

In the statistical analysis of time series, autoregressive–moving-average (ARMA) models provide a parsimonious description of a (weakly) stationary stochastic process in terms of two polynomials, one for the autoregression (AR) and the second for the moving average (MA). It predicts the future price from the prices of several past steps. Specifically, its prediction is made as a linear combination of the past prices subject to a Gaussian noise term.

A. Training set
We import ARMA model from statsmodels.tsa.arima model toolbox, and choose the parameters of p = 1 and q =2 from previous data analysis. We choose the bitcoin price dataset of every minute on the date of December 14,2018, and use it to train the ARMA model and get the prediction on this training set.

B. Testing set
Then we use the model in the previous part to make prediction of the next two hours on the date of December 15,2018. We put all the previous data in a list called history, and for every next minute, we use the ARMA model.fit to get the model, and then use model.forecast to get the next prediction. After the forecast, we add the true value of the current minute to the history list to help for the next prediction. Below is a comparison of the prediction label and true label.

C. Analysis
From the result, we can see the model fits well on short time prediction especially in one minute. However, it cannot predict on long term prediction because of the limit of linear stationary hypothesis. As a consequence, we implemented another machine learning method of LSTM model to realize long term prediction.

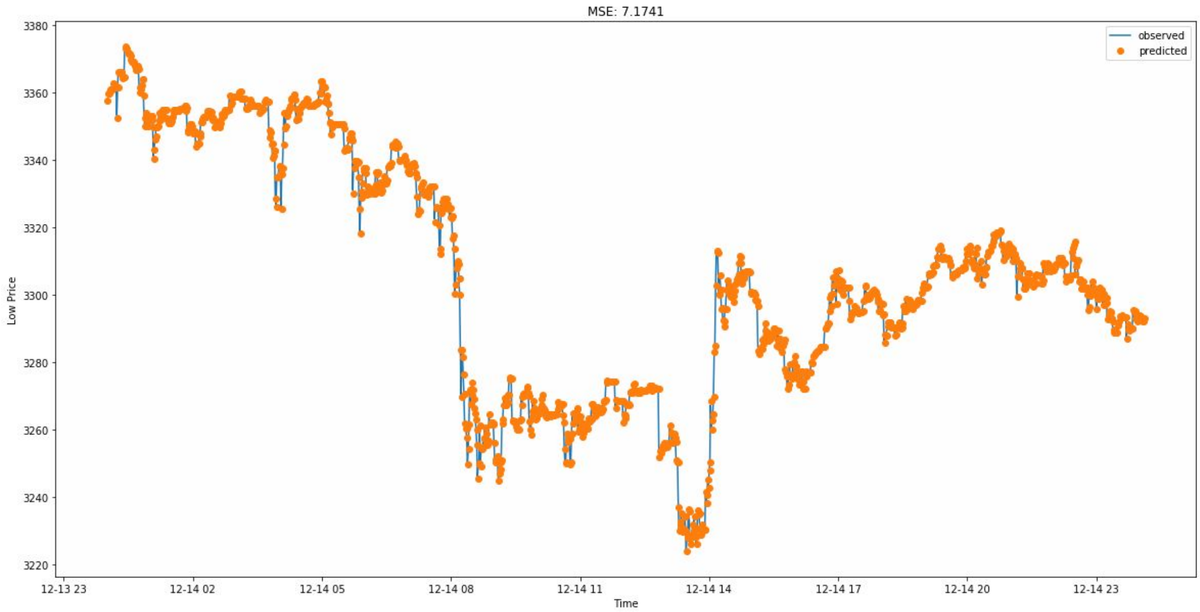

Figure 4: ARMA model on the training set

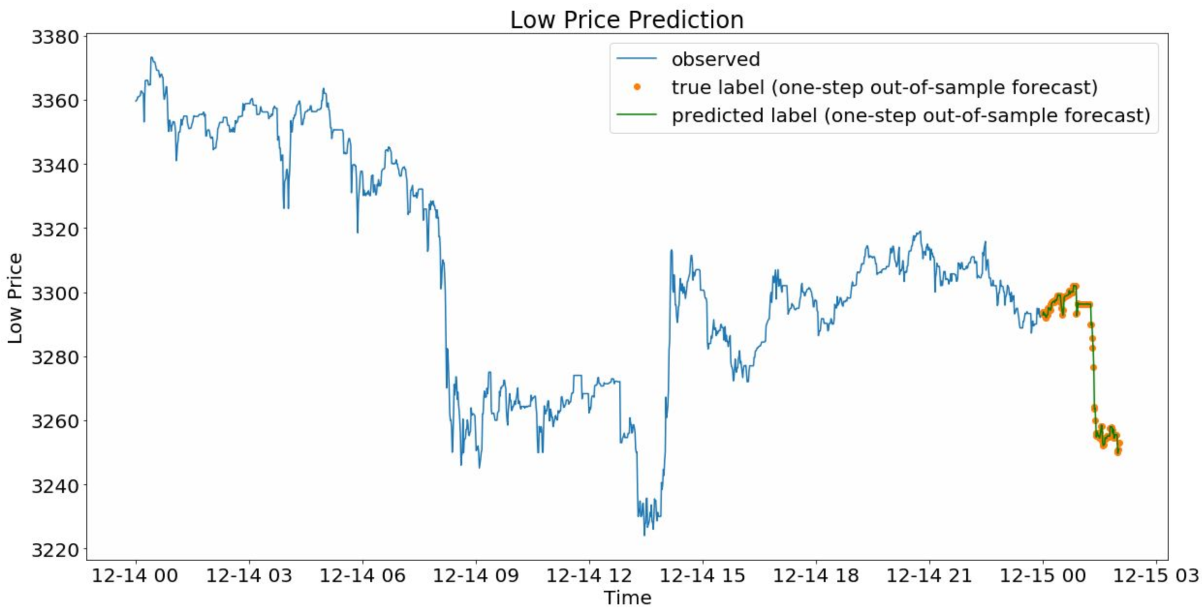

Figure5 : ARMA model on the training and testing set

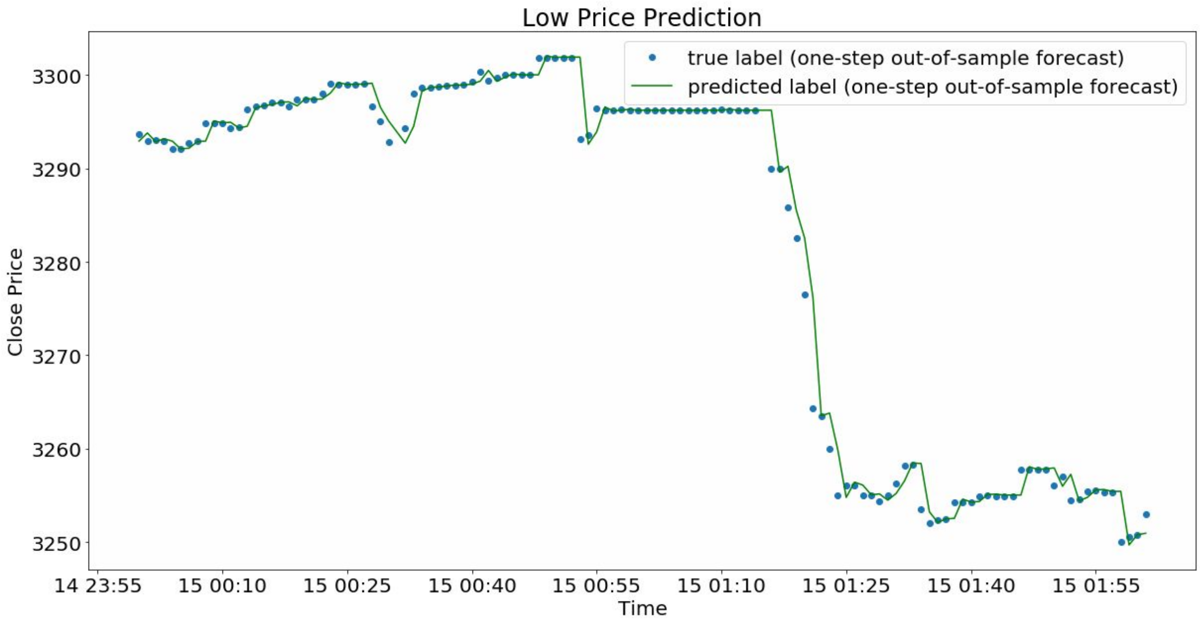

Figure 6: ARMA model on the testing set

LSTM Model

We're going to employ a Long Short Term Memory (LSTM) model; it's a particular type of deep learning model that is well suited to time series data. In this section, we present how we use LSTM model to predict Bitcoin price. LSTM(Long short-term memory) is an improved version of Recurrent Neural Network, which solves the long-term dependence of RNN. In the model, we use the prices and volumes of previous 7 days to predict the next day's price.

A. Data

Before defining the model, we first get one year's bitcoin prices and volumes using the bitfinex API, and then we preprocess data using pandas to get the 10 dimensional data
shown in Fig.

B. Model

The LSTM model has parameters including time lengths, input dimensions, layers, units. In our model, we set time lengths to be 7 to include the previous price and volume. We set input dimensions to be 2 to include both the volume and price. The output is one dimensional. Since we are building a simple LSTM model, we set the layers to be 4 including two dense networks and we use dropout to prevent overfitting. Finally, we compile the model using MSE lost function and adam optimizer. After one hundred epochs we get our model trained.

C. Analysis

From the result we can see that as time increases, the prediction get worse. Though we have are normalized data, the output is still bad, looking like this simple model overfit the training data. There are two main reasons. The first is that bitcoin prices are more than time series, the trend is changing from time to time. The second is that there are a lot of preprocessing need doing, like logarithm.

	Unnamed: 0	MTS	VOLUME	OPEN2	OPEN3	OPEN4	OPEN5	OPEN6	OPEN7	OPEN8	OPEN1
395	4	2018-12-11 16:00:00	20930.780792	-0.299557	-0.292092	-0.300637	-0.303519	-0.299067	-0.283974	-0.270738	-0.305082
396	3	2018-12-12 16:00:00	22248.786278	-0.305082	-0.299557	-0.292092	-0.300637	-0.303519	-0.299067	-0.283974	-0.298772
397	2	2018-12-13 16:00:00	28559.127026	-0.298772	-0.305082	-0.299557	-0.292092	-0.300637	-0.303519	-0.299067	-0.310161
398	1	2018-12-14 16:00:00	17889.937985	-0.310161	-0.298772	-0.305082	-0.299557	-0.292092	-0.300637	-0.303519	-0.314720
399	0	2018-12-15 16:00:00	6807.815932	-0.314720	-0.310161	-0.298772	-0.305082	-0.299557	-0.292092	-0.300637	-0.314682

Figure 7: Input Data Example

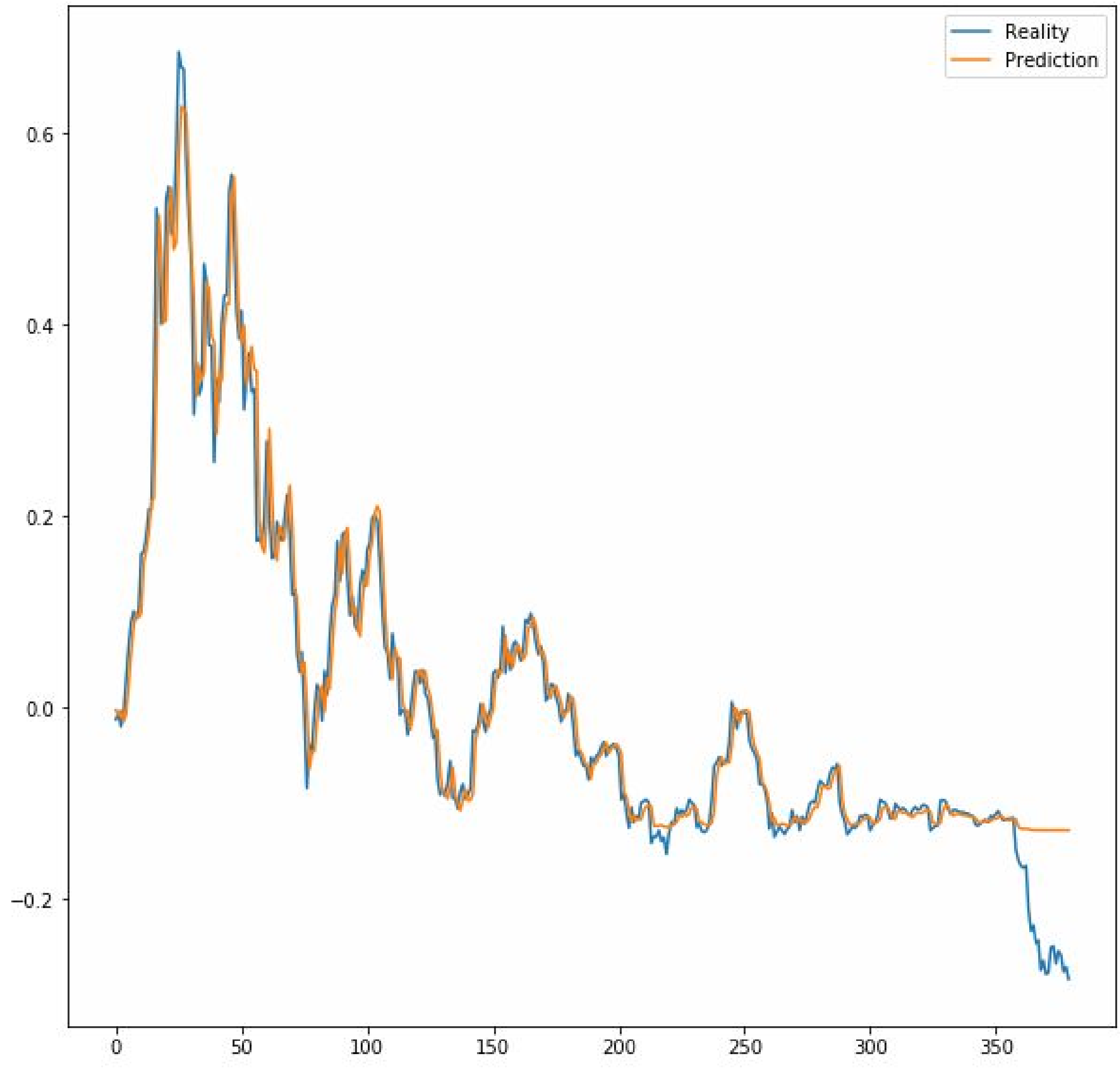

Figure 8: LSTM Model

Summary

In this report, we present the use of ARMA and LSTM models to make short and long term prediction of bitcoin price. LSTM performs better than ARMA, it does better on recognizing longer-term dependencies. However, we can tell from the result of two models that LSTM will need complex parameterization, which means LSTM cost lots of time. Besides, the shortcoming is that they both cannot predict the turning point of the prices, although both models can yield a similar prediction in term of MSE,. Therefore, we can do something to improve the model. Firstly, the bitcoin price is not stationary, some adjustments should be made to ensure that the p-value is below 0.01. Secondly, the LSTM only includes the historical data from the training set, which is. Reinforcement learning should be incorporated to update the model to make it be able to respond to the future changes.

Reference:

[1] Wikipedia contributors. Bitcoin [Internet]. Wikipedia, The Free Encyclopedia; 2018 Nov 25, 22:16 UTC [cited 2018 Nov 26]. Available from: https://en.wikipedia.org/w/index.php?title=Bitcoin&oldid=870603085

[2]BITCOIN PRICE - LATEST UPDATES: CRYPTOCURRENCY RECOVERS FROM EIGHT-MONTH LOW [cited 2018 12 16]. Available from: https://www.independent.co.uk/life-style/gadgets-and-tech/news/bitcoin-price-live-cryptocurrency-markets-ethereum-value-usd-latest-updates-a8491841.html