BITCOIN PRICE PREDICTION USING GCP AND DEEP LEARNING

Shubham Pawar

B.Tech Graduate, Dept of Computer Engineering & IT , VJTI College, Mumbai, Maharashtra, India.

Abstract --- Bitcoin is the first digital decentralized cryptocurrency that has shown a significant increase in market capitalization in recent years. The objective of this paper is to determine the predictable price direction of Bitcoin in USD by deep learning techniques and google cloud platform. Several algorithms of machine learning using supervised learning were explored to develop a prediction model and provide informative analysis of future market prices. Due to the difficulty of evaluating the exact nature of a Time Series(ARIMA) model, it is often very difficult to produce appropriate forecasts. Then I continued to implement Recurrent Neural Networks (RNN) with long short-term memory cells (LSTM). Thus, analyzed the time series model prediction of bitcoin prices with greater efficiency using long short-term memory (LSTM) techniques.

I. INTRODUCTION

Bitcoin price prediction has been an active area of research for a long time. Bitcoin, as a pioneer within the blockchain monetary renaissance, plays an overwhelming part in an entirely cryptocurrency market capitalization environment. Hence, it is the incredible interest of the machine learning and data mining community to be able to: (I) predict Bitcoin price changes (II) grant experiences to get it what drives the Bitcoin instability and way better assess related dangers in cryptocurrency domain.
People and big financial corporations are drawn to cryptocurrencies due to the transparency and anonymity that they offer to their users, in addition to their resistance to fraud because of the dispersed nature of the ledger statistics. Moreover, purchasing cryptocurrencies are promising in terms of making income, and ought to be of interest to buyers. In addition to familiarizing themselves with industry tendencies and political and financial information, they can utilize machine learning models to decide whether to buy or promote cryptocurrency. Hence, in this project, I used a GCP and deep learning to leverage machine learning technology to predict the real-time price of Bitcoin.

1.1. Bitcoin: It is a global cryptocurrency and online payment system that is highly stable and secure. It is Peer-to-peer value transfer and transaction protocol. Bitcoin transactions are verified by network nodes, published on a public ledger. Highly stable and secure Decentralized verification (Blockchains). “Bitcoin” is the unit of account on this ledger. Smallest unit: a satoshi representing 0.00000001 bitcoin, one hundred millionths of a bitcoin. In 2017, the Bitcoin price rose from $900 at the beginning of the year to nearly $20000 at the end of the year.

1.2. Price Prediction: Bitcoin's price varies similarly to a stock in another way. There are some algorithms used on stock market data for price prediction but the parameters affecting Bitcoin are distinctive. Therefore, it is essential to expect the price of Bitcoin in order that correct investment decisions can be made. The price of Bitcoin does not rely on business events or intervening government in contrast to the stock market. Hence, to expect the value it is essential to leverage machine learning technology to expect the rate of Bitcoin.

1.3. Problem Statements:

1. Bitcoin is the most complex cryptocurrency whose value changes every second.
2. Investing money for Bitcoin is more risky and less profitable.
3. Use ML and GCP for the analysis and prediction of Bitcoin price rate.

1.4. Project Objectives:

1. To predict bitcoin price with maximum efficiency using LSTM and ARIMA.
2. To compare between ARIMA and LSTM to find which is the best efficient algorithm for predicting bitcoin price.
3. To ensure less risk and more profit for investors.

II. METHODOLOGY

ANALYSIS AND DESIGN

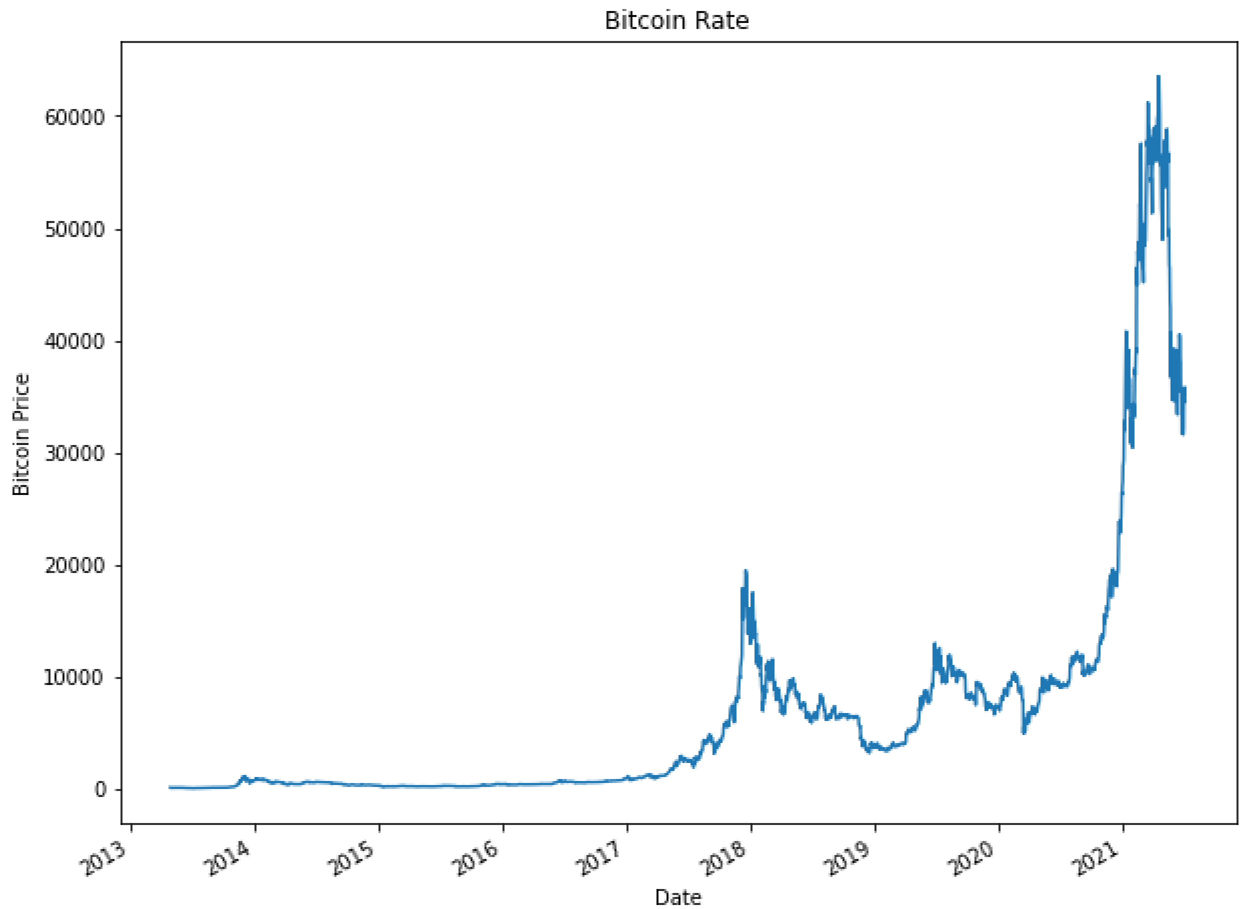
Collection of Datasets:
For the better prediction of price I have used two datasets. First one is from Kaggle which contains price history on a daily basis from April 28, 2013 to Feb 27, 2021. And the second one is from "https://in.investing.com/crypto/bitcoin/historical-data" which contains price history from Feb 28, 2021 to June 30, 2021. The datasets have the historical price information of Bitcoin by market capitalization.

- Date : date of observation
- Open : Opening price on the given day
- High : Highest price on the given day
- Low : Lowest price on the given day
- Close : Closing price on the given day
- Volume : Volume of transactions on the given day
- Market Cap : Market capitalization in USD

Merging of Datasets: The goal is to predict Bitcoin price from July 1, 2021 to June 30, 2022. Hence in this phase, two different datasets were combined into a single dataset so that we can predict the price for the next 365 days.

Data Ingestion in the BIGQUERY: BigQuery is a fully-managed, serverless data warehouse that enables scalable analysis over petabytes of data. It is a Platform as a Service that supports querying using ANSI SQL. It also has built-in machine learning capabilities.

Data Analysis using BIGQUERY and DATA STUDIO: BigQuery is a petabyte-scale analytics data warehouse that you can use to run SQL queries over vast amounts of data in near real-time. Data visualization tools can help you make sense of your BigQuery data and help you analyze the data interactively. You can use Google Data Studio, a visualization tool to help you identify trends, respond to them, and make predictions using your data.



The above figure depicts that higher fluctuation in the bitcoin price occurs during 2017 to 2018 and again extreme price growth during 2020 to 2021.

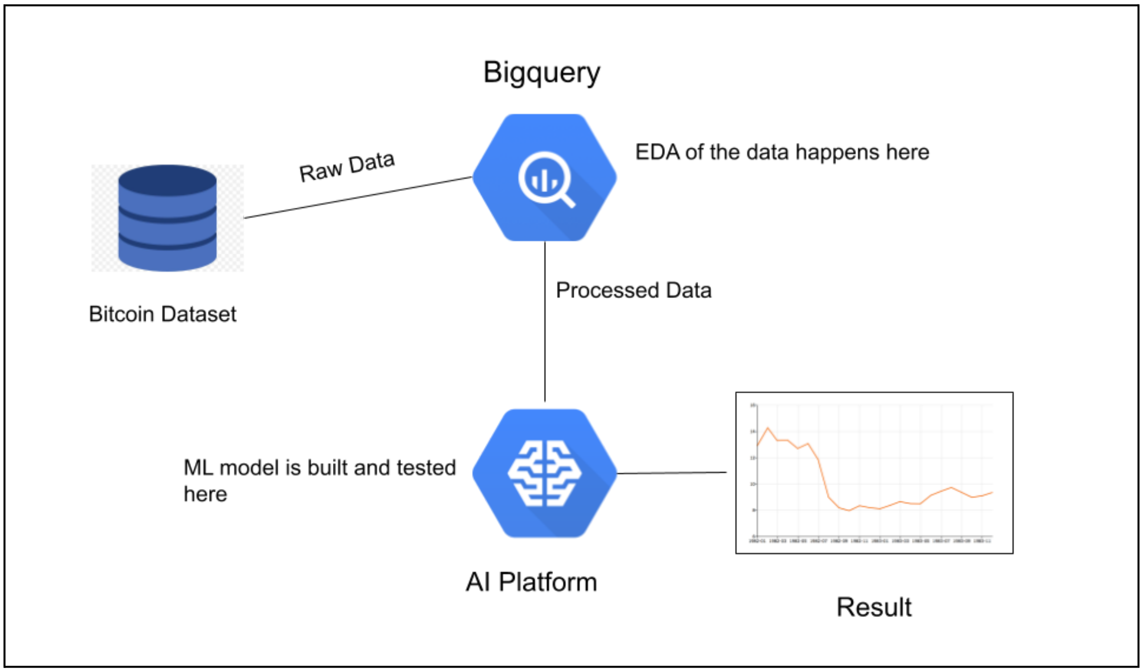
Data Cleaning: It is the process of preparing data for analysis by removing or modifying data that is incorrect, incomplete, irrelevant, duplicated, or improperly formatted.

Data Normalization: We need to normalize inputs in neural networks and other data mining models, otherwise, the network will perform poorly. Normalization is carried out in order to have the same range of values for each RNN model input. This can guarantee stable weight and partiality convergence. Normalization here uses the MinMaxScalar Package, after normalization, data is plotted using matplot libraries and the trend is seen to check the fluctuations in the price.

Splitting of Data: We divide the data into training and test data using the Scikit Library. The data for training is divided into 65 percent and 35 percent for testing.

IMPLEMENTATION

Following figure shows the complete procedure to predict the Bitcoin price rate. The raw data is ingested in the Bigquery. Here we do EDA to get better insights of the data, also we perform some fundamental preprocessing. Then the processed data is used to train and test the machine learning model. Here we use the cloud environment provided by GCP to build our model.



LSTM Modeling

The long short-term memory network or LSTM addresses the common problem of disappearing gradients in the recurrent neural network. This is a type of recurrent neural network that is used in profound learning, as very large architectures can be trained. LSTM enables the network to learn more about many time steps by maintaining a more-steady error. This enables the network to learn long-term trust. LSTM cell contains forget and remember gates that allow the cell to decide which information to block or transmit based on its strength and importance . As a result, weak signals that prevent the gradient from disappearing can be blocked.
The function below converts the series to supervised data. Two models are available in Keras. One is a sequential model that is suitable for predicting time series and the other is used with a functional API. The dense layer is used with input shapes as the output layer. The optimizer function used is 'adam' which has a learning rate of 0.01 with the mean absolute error as the loss function. The loss method used means absolute error.

ARIMA Modeling

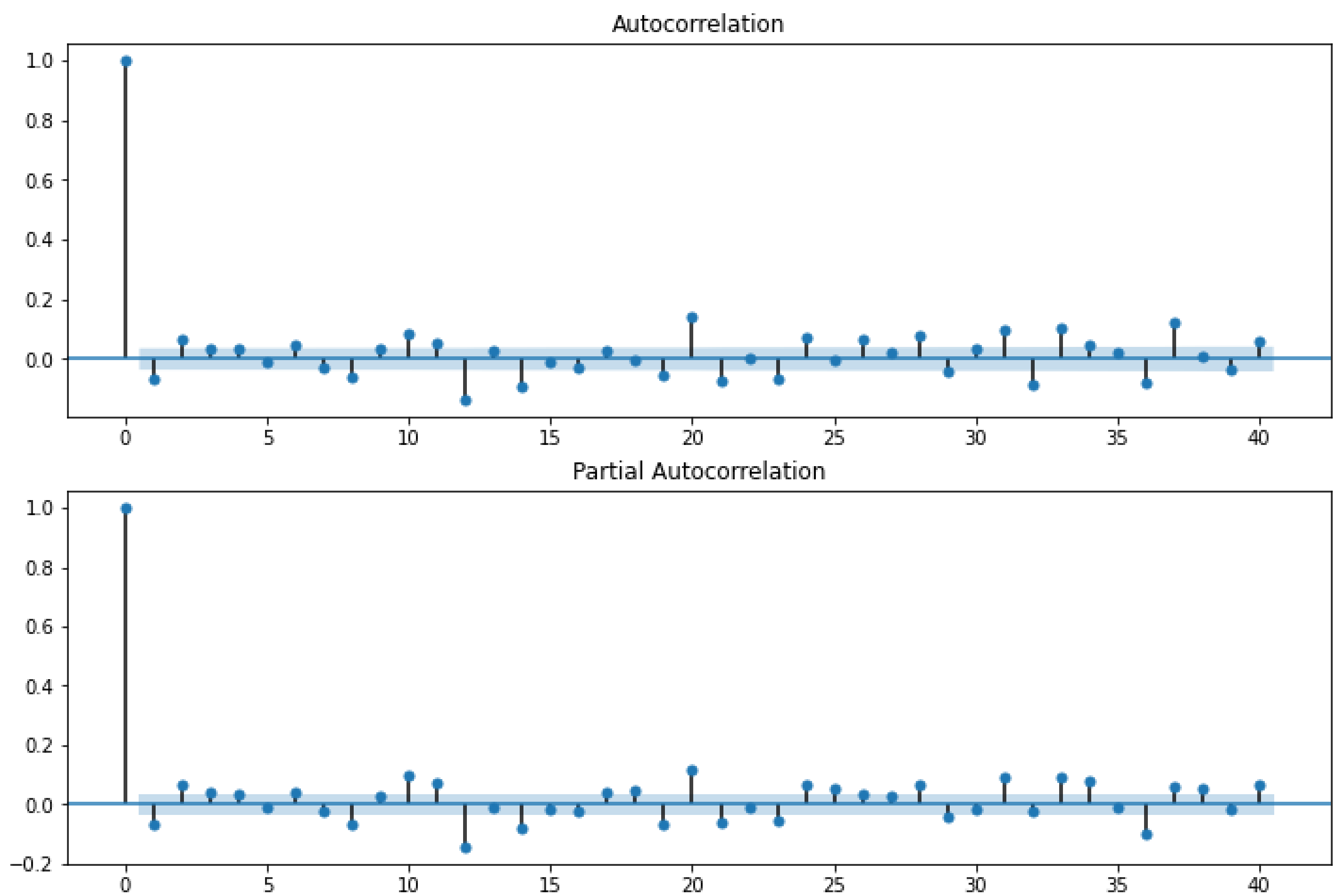
ARIMA model uses and decomposes historical data into Autoregressive (AR) Indicates weighted average movement over past observations, Integrated (I) Indicates linear trends or polynomial trends and Moving Average (MA) Indicates weighted average movement over past mistakes. Therefore, the model has three parameters AR(p), I(d) and MA(q) all combined to form ARIMA (p, d, q) where p= autocorrelation order d= integration order (differentiation) q= moving averages. A non - seasonal stationary time series may be modeled as a combination of past values and errors known as ARIMA (p, d, q) or expressed as ARIMA (p, d, q). To fit an ARIMA model that assumes stationary characteristics, we must use our data to determine the three parameters: p;d;q. p corresponds to the autoregressive component, q corresponds to the moving average and d corresponds to the number of nonseasonal differences needed for stationarity.

III. RESULT

From the experiments, the result shows that machine learning models (LSTM) take much longer to compile because of their complex calculations than traditional models(ARIMA).

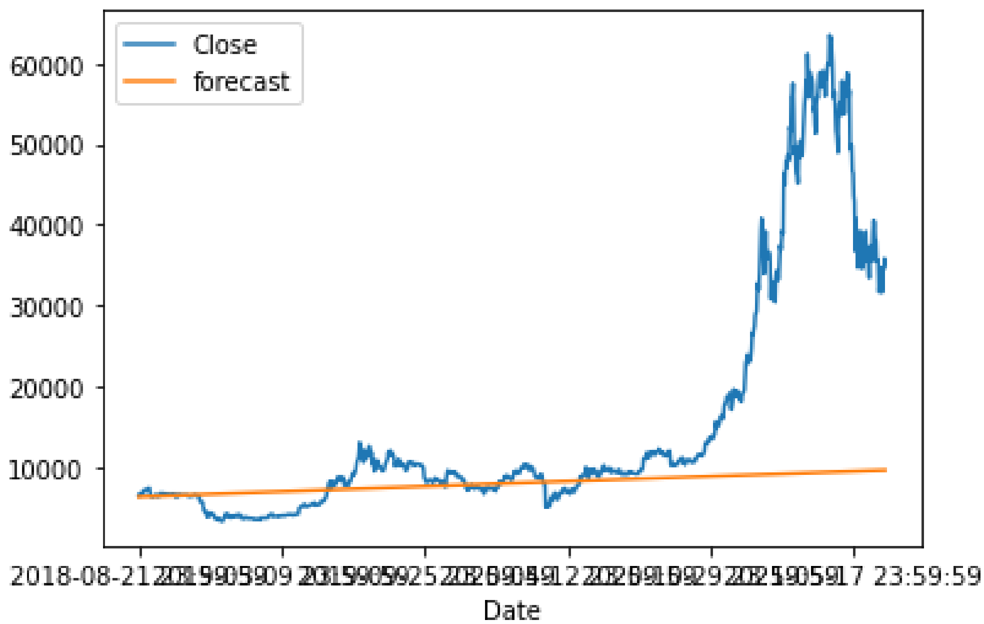
ARIMA Model :

Identification of an AR model is often best done with the PACF. Identification of an MA model is often best done with the ACF.



So the graph shows, for non-seasonal data order should be p=1, d=1, q=1.

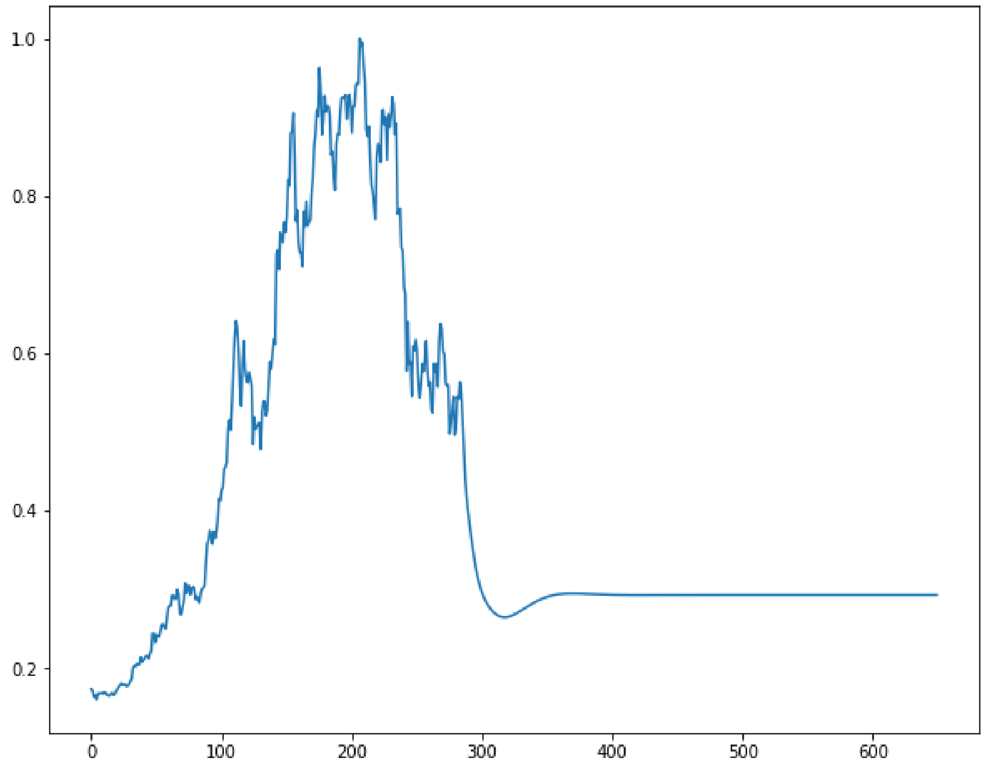
Following graph shows that the prediction result is so poor with the ARIMA MODEL.



LSTM Model :

In the following fig blue line represents the actual data, orange line represents the prediction result for training data and green line represents prediction result for the testing data.

Following graph shows the prediction for the next 365 days i.e. from July 1, 2021 to June 30, 2022.



Since the performance of the model in LSTM is far better than ARIMA, the machine learning algorithm (Long Short Memory) is best suited for Bitcoin forecasting. LSTM is generally suitable for predicting higher fluctuations in the time series data.

IV. CONCLUSION AND FUTURE SCOPE

This study focuses on the Bitcoin closing price for the development of the predictive model. The prediction is limited to previous data. The model developed using LSTM is more accurate than the traditional models that demonstrate a deep learning model. In our case, LSTM (Long Short-Term Memory) is obviously an effective learner on training data than ARIMA, with the LSTM more capable of recognizing long-term dependencies.

This study uses the daily price fluctuations of Bitcoin, to further improve the model's predictability we can consider data with hourly price fluctuations in the future. This paper consists only of comparing ARIMA with LSTM. The result would be confirmed by comparing more machine learning models in the future.
We can apply sentiment analysis and supervised machine learning principles together to the extracted tweets from Twitter and Reddit posts, and we can analyze the correlation between bitcoin price movements and sentiments in tweets. This will further improve the model.