Bitcoin closing price prediction

Name:	**Sharanyak Podder**
Registration No./Roll No.:	19287
Institute/University Name:	IISER Bhopal
Program/Stream:	DSE
Problem Release date:	February 02, 2022
Date of Submission:	April 24,2022

1 Introduction

*B*itcoin is a virtual currency created by Satoshi Nakamoto in 2008. Because of its revolutionary properties, transparency, simplicity, and growing acceptance, Bitcoin has become increasingly popular and used for financial transactions worldwide. The rising use of machine learning techniques to predict time series and the use of cryptocurrencies as financial instruments prompted the current study to look for more accurate Bitcoin price predictions. In this paper, we examine the short-term prediction of the Bitcoin market using several machine learning approaches and a wide range of potential predictive market variables. 1.

Figure 1: The Change in Price of Bitcoin Over Time

2 Problem Statement

On the given dataset try to implement supervised learning algorithms to predict the closing price of the Bitcoin based on the opening price of the Bitcoin.

3 Methods

Data : We've given a dataset with six features and with 749 rows in the training dataset and 250 rows in the test dataset. The data ranges from April 17, 2015, to January 6, 2018, and includes columns for Date, Timestamp(ts), Weighted Price(wp), Highest Price(hi), Lowest Price(lo), and Bitcoin Opening Price(open). The labels are the date and the Bitcoin closing price(close) for that day.

Data preprocessing : First we merged the dataframes of train and label to include the closing price in our dataset.Then we sort our dataset serially by date. The 'date' feature is categorical so we convert it to datetime object. Then we checked for null values in our dataset.In this project, I have implemented a few predefined python libraries which help in data visualization and can help you understand the important features which are required by the system. With the help of data visualization libraries, we can see the correlation between features and pinpoint the ones which we require. 2 We can observe that all the features except timestamp(ts) are perfectly positively linearly correlated i.e. correlation = 1 so we can conclude that ts has not that importance other than wp, hi, lo and open. we have visualized monthwise and yearly comparison between open and closing price of Bitcoin. 3

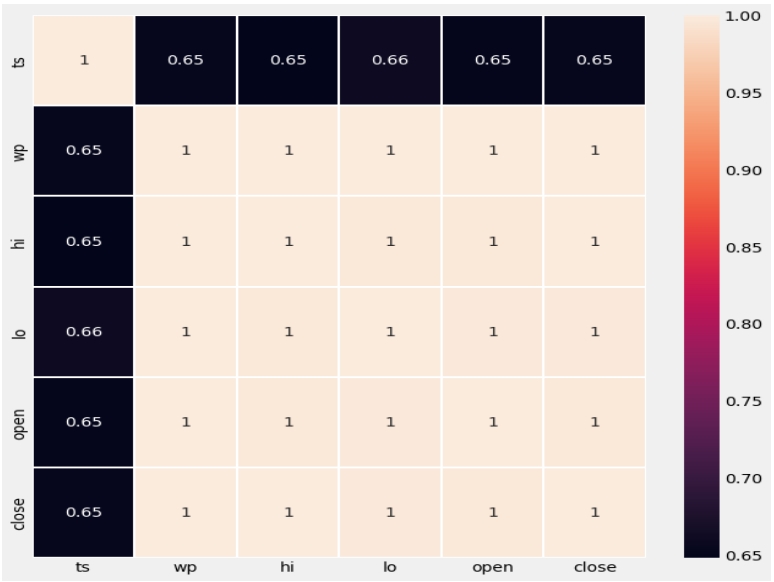

Figure 2: The Change in Price of Bitcoin Over Time

Data Preparation : When variables that are measured in different scales it does not contribute equally in model fitting which will lead to model learned function to create a bias. Thus, standardization or normalization of data is very much essential for better accuracy and result. When working with a Machine Learning model or Deep Learning models where we require back propagation to be more stable and even faster, proper scaling of data is necessary.

Algorithms Implemented : ARIMA is a class of models that captures temporal structures in time series data. ARIMA is a linear regression based forecasting approach. Therefore it is best for forecasting one-step out-of-sample forecast. Here, the algorithm developed performs multi-step out-of-sample forecast with re-estimation, i.e., each time the model is re-fitted to build the best estimation model.

Unlike modeling using regressions, in time series datasets there is a sequence of dependence among the input variables.Recurrent Neural Networks are very powerful in handling the dependency among the input variables. LSTM is a type of Recurrent Neural Network (RNN) that can hold and learn from long sequence of observations.

XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. In prediction problems involving unstructured data (images, text, etc.) artificial neural networks tend to outperform all other algorithms or frameworks. However, when it comes to small-to-medium structured/tabular data, decision tree based algorithms are considered best-in-class. Our

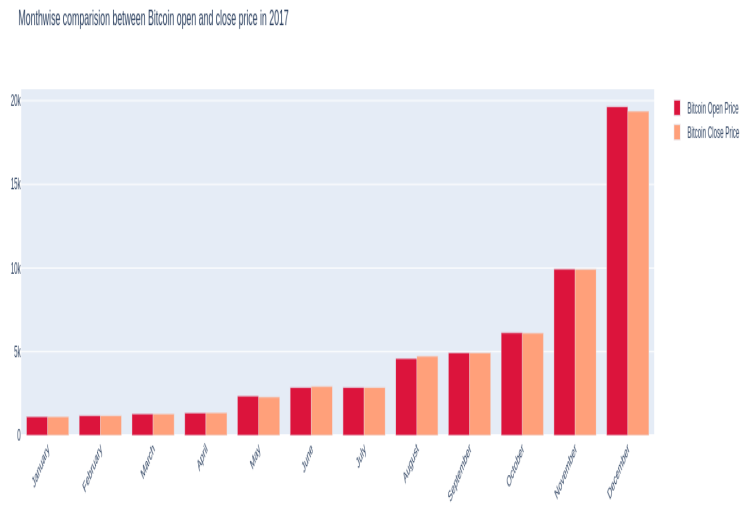

Figure 3: The Change in Price of Bitcoin Over Time

given dataset is small enough thus we applied XGBoost.

4 Evaluation Criteria

The prediction models are evaluated and analyzed regarding various aspects. we compare the models on a prediction level.We compare the forecasts of our prediction models based on the predictive accuracy on the test set. 4 5

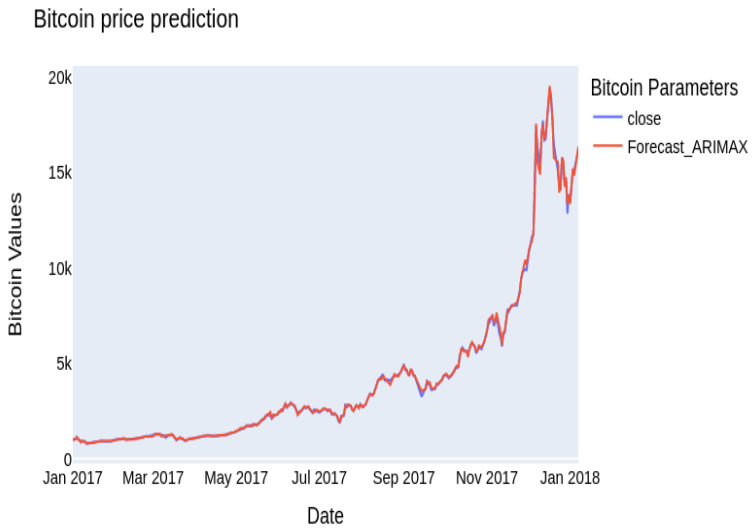

Figure 4: Forecasting ARIMA

5 Analysis of Results

LSTM accuracy scores :

Train data RMSE	7.409199631113801
Train data MSE	54.896239173696884
Train data MAE	4.144708195753732

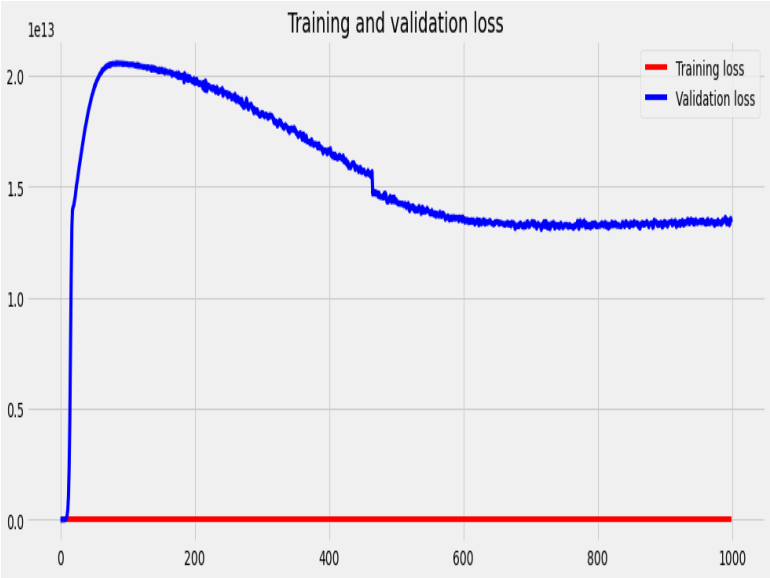

Figure 5: LSTM model training and validation loss

6 Discussions and Conclusion

We have seen that ARIMA is best fitted model. So we are selecting ARIMA for our baseline model and using this model to predict the closing price. Because the crypto market is volatile and influenced by social media and other external factors, data sets cannot be the only reason for forecasting. As technology advances, new data can be collected, analyzed, and practiced, resulting in better results for this experiment.

Future Scope : 1.To work on a better User Interface so that people can access these data easily and effortlessly. 2.Implementing IOT model for smart automatic analysis. 6

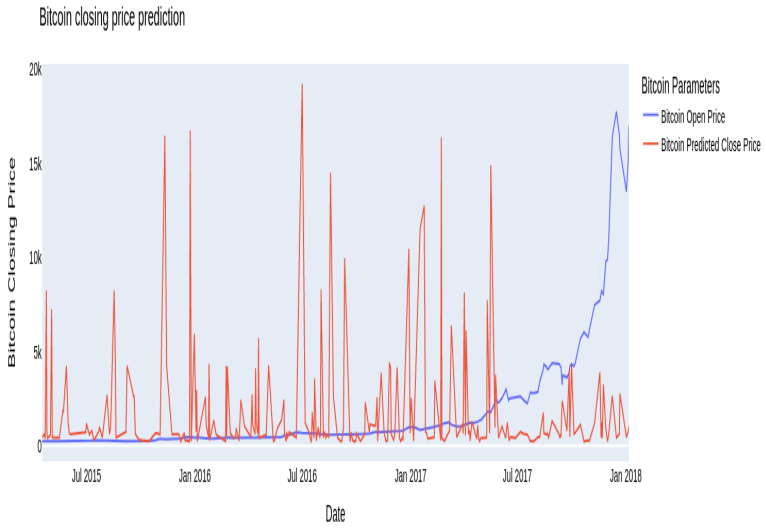

Figure 6: Predicting the closing price of Bitcoin