Pattern Recognition and Machine Learning

BITCOIN PRICE PREDICTION

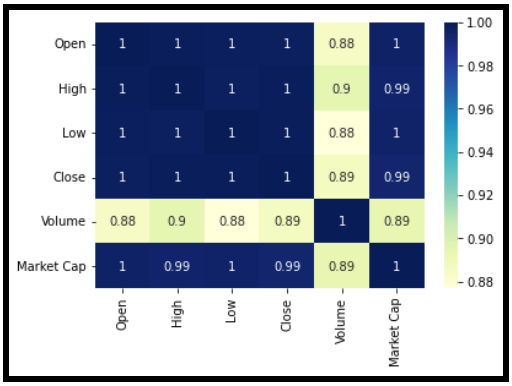
SARVESH KULKARNI (B20EE059)

Introduction

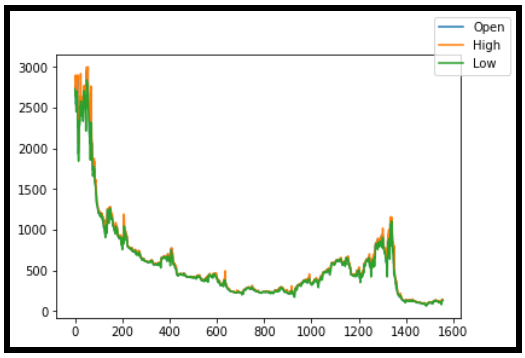
Bitcoin is a decentralized digital currency, without a central bank or single administrator, that can be sent from user to user on the peer-to-peer bitcoin network without the need for intermediaries. Transactions are verified by network nodes through cryptography and recorded in a public distributed ledger called a blockchain.[0]

Dataset / Visualization and Preprocessing

The dataset has 1556 rows with 7 columns, namely Date, Open, High, Low, Close, Volume, and Market Cap. Certain column types were changed from object to int. The correlation map of features - after dropping unnecessary ones - is as shown:



It is clear that the features are heavily correlated. It can be confirmed by plotting all Open, High, and Low together giving us the plot:



Hence, the features Open, High and Low are chosen to be dropped.

Methodology

Several models were experimented with for forecasting and 2 models - LGBMRegressor and Linear Regressor - had the best and nearly similar results in testing data. The method used is discussed below.

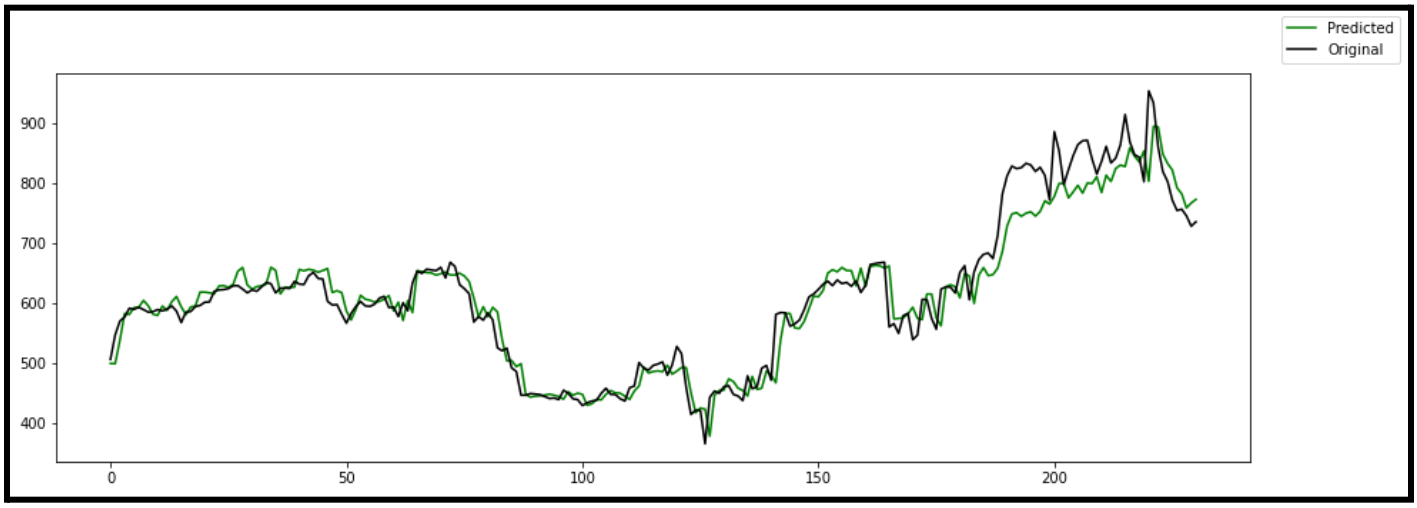
Each price is treated as a prediction, where the prices for the same parameter for the last n (here, 32) days are treated as corresponding features. A new dataset is created in such manner. Models are trained on the dataset that can predict the future price for a parameter, given the price for the same parameter for the n days.

LGBM Regression

LightGBM extends the gradient boosting algorithm by adding a type of automatic feature selection as well as focusing on boosting examples with larger gradients. The model was fitted to the training dataset and then predicted on the training dataset created.

A dataset was created as discussed above and trained on the LGBM Regression model. The predictions were recorded and compared with the original data.

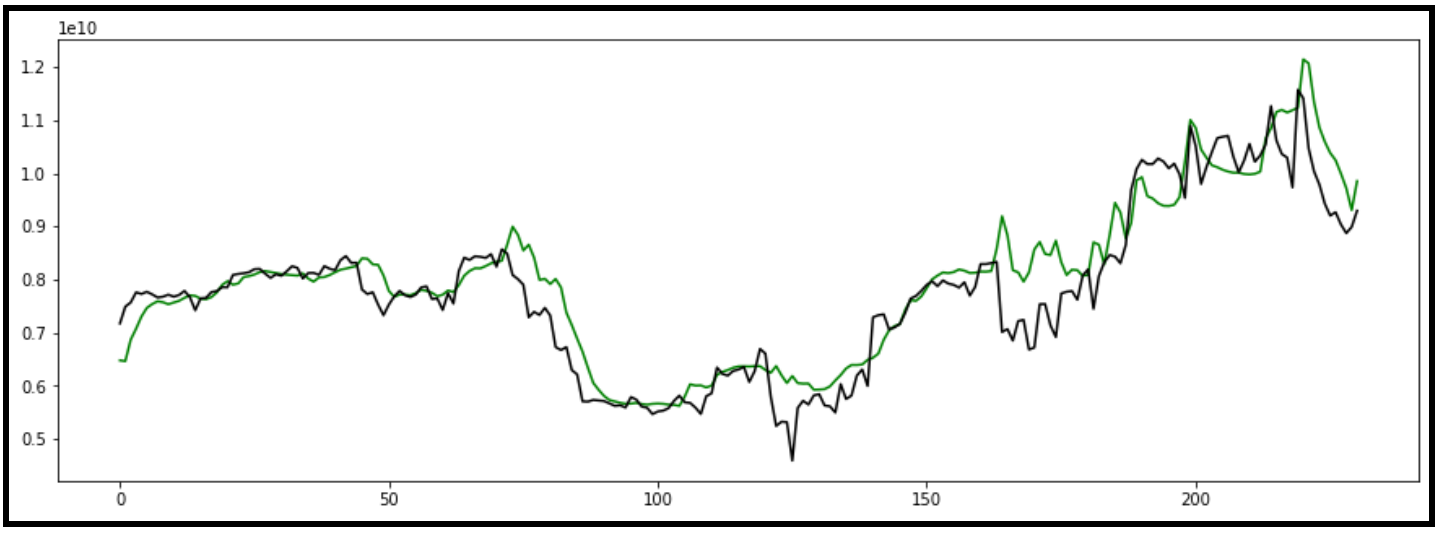
The forecast for Close prices is plotted along with the originally given values:



It can be seen that the forecasted price matches significantly with the provided data. The R-squared value between them is reported to be 0.927.

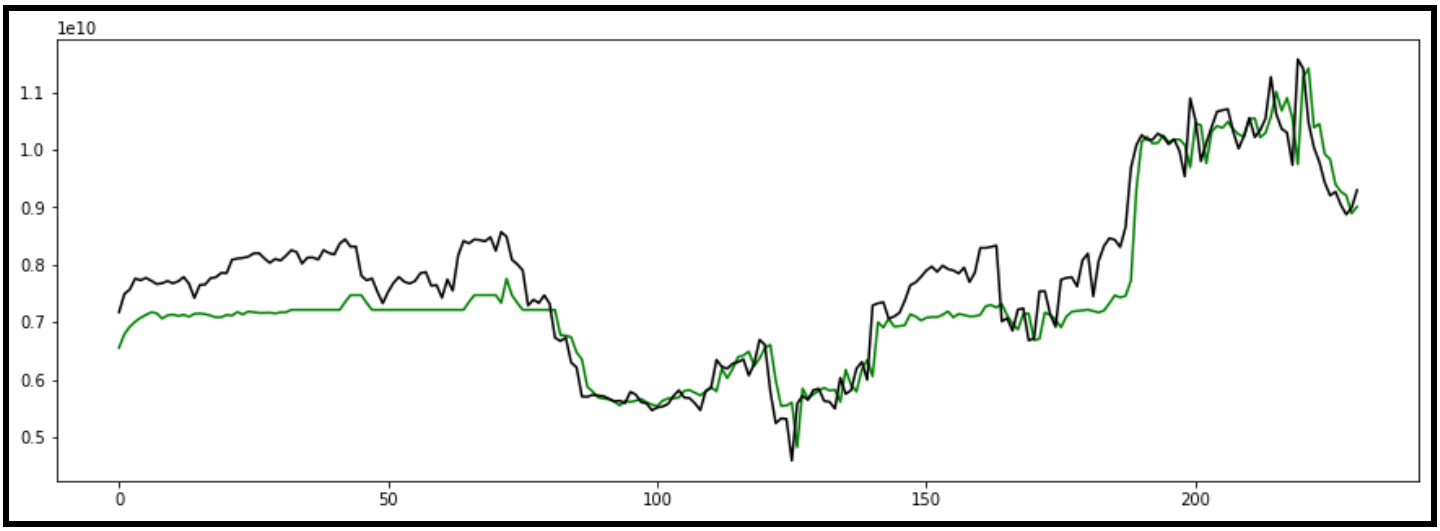
Now, forecasting Market cap is done in 2 ways. One is to directly use historical data of the same, and the other is to forecast the Close and Volume feature and hence predict the corresponding Market cap.

The forecast for Market Cap values (using historical data of self only) is plotted along with the originally given values:

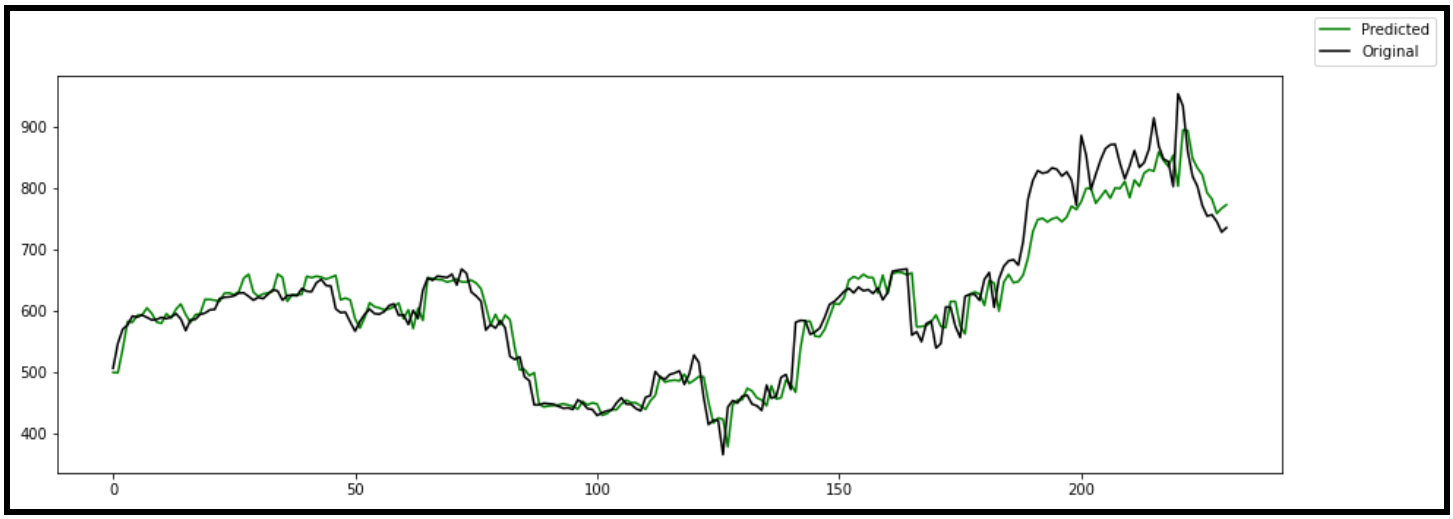


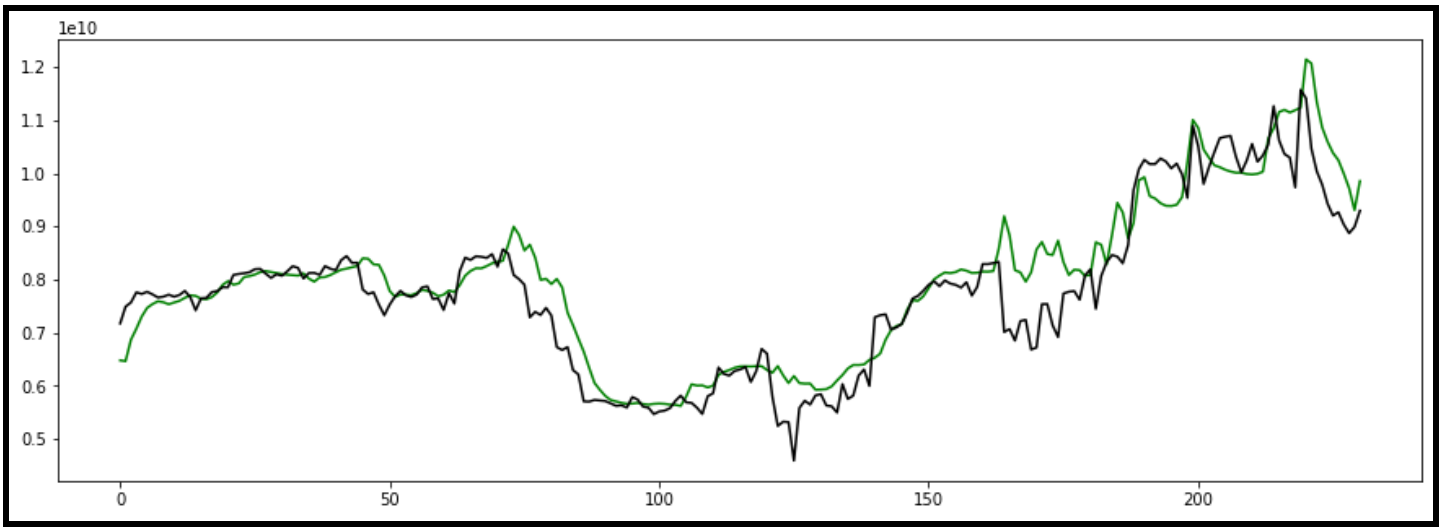
The R-squared between them is reported to be 0.821.

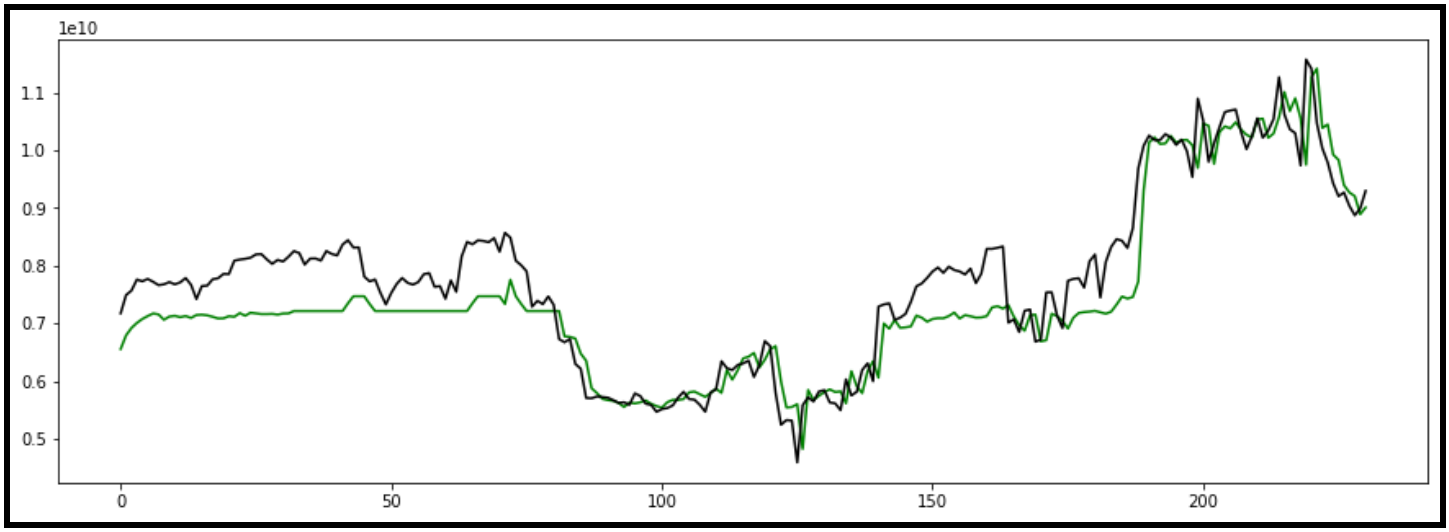
The forecast for Market Cap values (based on the prediction from corresponding forecasted Close and Volume features) is plotted along with the originally given values:



The corresponding plots for Linear Regressor are shown below:




1e10
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0
50
100
150
200


1e10
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0
50
100
150
200