15 may-2022
machine learning

Bitcoin Price Prediction

Group members:

Dania Alqarni, ID: 1911223
Lujain Alharthi, ID: 1915281
Hibah Sulays, ID:1915408
Atheer Alsahli ,ID:1911154

Contents

Introduction

Bitcoin is decentralized digital money that may be sent directly from one user to another on the bitcoin network without the need for a central administrator. Network nodes use cryptography to validate transactions, which are then recorded in a public distributed ledger known as a blockchain, which also tracks ownership, prevents tampering with transaction records, and prevents double-.spending

The country's financial development was triggered by massive trade shifts. Changes in the presentation of computerized monetary standards such as bitcoin (BTC) from a cash-based exchanging framework, for example, are crucial. The underlying cost of BTC was $0.0008 in Fig. 1.1, and the long-term cost of BTC has climbed sharply to $46,434.40

.

There are many different ways to make money with based on the value and complexity involved, predicting cryptocurrency is a challenging and exciting task for traders, investors, and researchers. The current research focuses on forecasting the future of the cryptocurrency 'Bitcoin.' The suggested prediction model for learning and prediction, the model uses a feature matrix formed on technical analysis as input the proposed model has been tested in the field.

The impact of Fed Funds rates on Bitcoin price dynamics is investigated Bitcoin prices reach an all-time high of almost $60,000 during the 2020–2021 bull run. The trend of businesses adding Bitcoins to their balance sheets has sparked widespread interest in Bitcoin as a secure private payment method. .While it had a nominal worth

The impact of Fed Funds rates on Bitcoin price dynamics is investigated Bitcoin prices reach an all-time high of almost $60,000 during the 2021 bull run. The trend of businesses adding Bitcoins to their balance sheets has sparked widespread interest in Bitcoin as a secure private payment method. While it had a nominal worth.

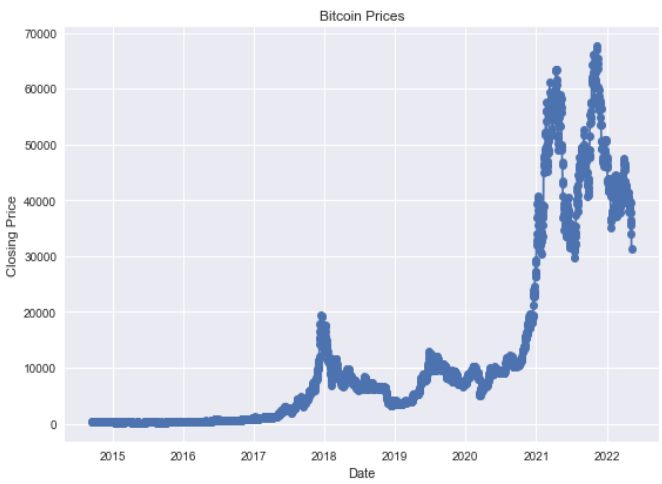


Fig1.1 We can notice a huge spike in the year 2021 and 2022,we may consider COVID-19 pandemic.

scenario and identify pre-existing sources of data.

Given the importance of electronic currencies and their current entry into simple life financial transactions such as buying bread and up to shares, we found a goal, which is to analyze the historical data set for the encrypted currency Bitcoin, as it is one of the highest traded and price. The data used is freely available on the Keggle website under title of the Top 100 Historical Datasets for Cryptocurrencies uploaded during May 2020

Develop and clarify objective

Can machine algorithms forecast the future price of bitcoin?
We believe that digital currencies (Cryptocurrencies) in general, and Bitcoin in particular, are the way of the future. We are working on a model that can learn and forecast the future market value of a currency, allowing us to determine if buying or exchanging that currency is in the buyer's best interests. To achieve the optimal result, an integrated analytical plan will be used.

The set of libraries used

NumPy	**Pandas**	**Matplotlib**	**RandomizedSearchCV**
sklearn	Keras	IPython	StandardScaler
	seaborn		RandomForestRegressor
mean_squared_error			
LinearRegression			
linear_model			

Dataset

This dataset was downloaded from Kaggle at https://www.kaggle.com/datasets/kaushiksuresh147/top-10-cryptocurrencies-historical-dataset

We needed to find a trading dataset that included a Market value at opening and closing in addition to the volume since we wanted to see if purchasing or .exchanging money was best able.
The dataset that is best suited to our research. Table1.1 shows the 7 columns in this .dataset

Name of attribute	Data Type
Date	date
Open	float
High	float
Low	float
Close	float
Volume	float
Currency	String ---constant value(USD)

Table1.1

approach

Our approach is to experiment to find the optimum solution. To begin, we looked for other people's experiences and analyses in same domain to find the most important algorithms for analysing this type of data. The four best algorithms for analyzing Cryptocurrencies(digital currencies) were discovered to be

- Linear Regression(simple , Multiple): Linear Regression is an approach for modelling the relationship between scalar dependent variable Y and one or more explanatory variables (or independent variables) X."
  useful measure for technical and quantitative analysis in financial markets.

- Neural Networks: is a group of algorithms that attempt to recognize underlying relationships in a batch of data using a method that mimics how the human brain works. In this context, neural networks are systems of neurons that might be organic or artificial.

- Random forest (Classifier , regressor): is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model also estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.

  The Working process can be explained in the below steps
  Step-1: Select random K data points from the training set.
  Step-2: Build the decision trees associated with the selected data points (Subsets).
  Step-3: Choose the number N for decision trees that you want to build.
  Step-4: Repeat Step 1 & 2.
  Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

  Random Forest Regression is a supervised learning algorithm that uses ensemble learning method for regression. Ensemble learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model.

  Random Forest is a classifier that contains several decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

As a result, all of these algorithms will be applied, and the results will be compared and the accuracy observed in order to identify which one of them is the best and apply it for future predictions.

How is the data divided between X and Y? There are two types that have been tested

Name of attribute	Data Type
Date	index
Open	X
High	X
Low	X
Close	X
Volume	Y

Name of attribute	Data Type
Date	X
Open	X
High	X
Low	X
Close	Y
Volume	X

Dataset Distribution:

- Closing Price

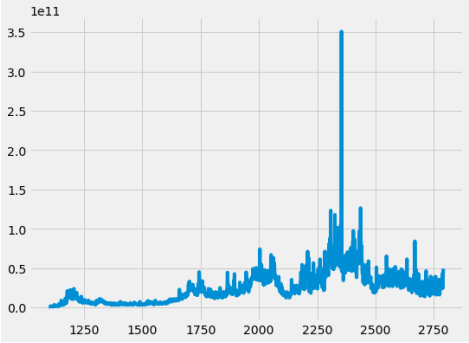
An adjusted closing price is a stock's closing price on any given day of trading that has been amended to include any distributions and corporate actions that occurred at any time prior to the next day's open .

- Opening price

The opening price is the value that each share has the opening price gives a good indication of where the stock will move during the day. Since the Stock exchange can be likened with an auction market i.e. buyers and sellers meet to make deals with the highest bidder, the opening price does not have to be the same as the last day's closing price.

- Volume

Volume is one of the most basic and beneficial concepts to understand
when trading stocks. Volume is defined as, "the number of shares or
contracts traded in a security or an entire market during a given period.



Algorithms result

1. Linear Regression

1.1 Simple linear Regression

In this case study we choose linear regression for our analysis. First, we divide the data into two parts of training and testing. We'll train a simple linear regression model using a 10-day exponential moving average (EMA) as a predictor for the closing price which is commonly used technical indicators by Investors to predict the movements of stocks which helps them to take the right decision in terms of to buy, sell, or hold. After that analyze the accuracy of our model, plot the results.

The Command used for add our technical indicators we'll be using the pandas_ta library. To get started, let's add an exponential moving average (EMA) to our data:

```
import pandas_ta
# Add EMA to dataframe by appending
# Note: pandas_ta integrates seamlessly into
# our existing dataframe
#This is our newly-calculated value representing the
# exponential moving average calculated over a 10-day period.

df.ta.ema(close='Close', length=10, append=True)
```

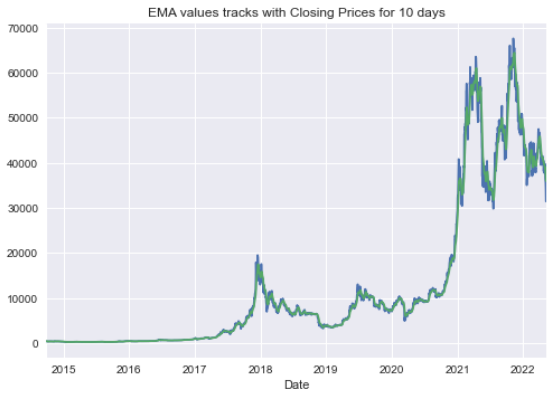
As evident from the printouts, we now have a new column in our data titled "EMA_10." This is our newly-calculated value representing the exponential moving average calculated over a 10-day period.

EMA_10
413.122
406.587
400.929
398.386
395.7
391.95
386.052
375.655
365.628
359.165
354.987

Now we've trained a model on 80% of historical pricing data using the Closing value and the Exponential Moving Average for a 10-day trading period. Our goal was to develop a model that can use the EMA of any given day (dependent on pricing from the previous 9 days) and accurately predict that day's closing price .

```
# Split data into testing and training sets
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(df[['Close']], df[['EMA_10']], test_size=.2)

from sklearn.linear_model import LinearRegression

# Create Regression Model
model = LinearRegression()

# Train the model
model.fit(X_train, y_train)

# Use model to make predictions
y_pred = model.predict(X_test)
```

The results of the prediction are shown below, whereas the blue line shows the actual close price and the green line shows the predicted close price of Bitcoin based on the EMA of 10 days.
With Accuracy = 94.74 %



And this result can professionally ensure that the prediction system by using linear regression with EMA can make an accurate predictions for Bitcoin prices based on a given number of days.

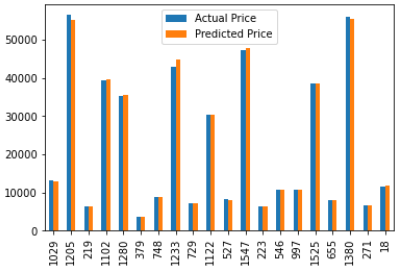
1.2 Multiple linear Regression

First in Multiple Linear regression, we divided the data into two parts to design the relationship between four independent variables and 1 dependent variables. Splitting the dataset so it can be treated, divided into training and testing set with 25%.

As shown, we have chosen Open, High, Low and Volume as an independent variables and the dependent variable to be predicted is Close price. As we mentioned above, our goal of this study is to propose an effective prediction system to help the investors to make the best decision.
import Linear Regression model, and create another variable is "regression" to control lineage. After applying the model, we have got the results:

Actual Price	Predicted Price
13075.2	13023.7
56473	55294.7
6305.8	6380.88
39371	39624.9
35287.8	35519.8
3630.68	3647.52
8706.25	8914.82
42909.4	44765.6
7193.6	7223.07
30432.5	30382.4
8145.86	8107.15

Plot the result to check and compare the actual and the predicted price for a randomly 20 row, we can see a minor difference between them in the below graph, which represents a good results of the model prediction.



The Accuracy is = 98.89%, which means the model is very well predicting and it can give the investors a trustable results.

2. Recurrent Neural Networks.

We have created a new data frame with only Close column then compute the number of row to train the model on about 80% of the dataset so we have 1272.

Scale the data because in practice it is nearly advantageous to apply pre-processing transformations scaling or normalization to the input data before its presented to a neural network. And help the model.
Scaler fit transform compute the min and max values to be used for scaling and they transform the data based on these two values, The range will be 0 to 1 inclusive.
Create the training dataset and the scale training dataset and testing
X_train will contain 100 values those will be indexed from position 0 to position 99 and then the y_train dataset for the first pass through will contain the 101st value which will be at position 100 that we want to predict as shown:

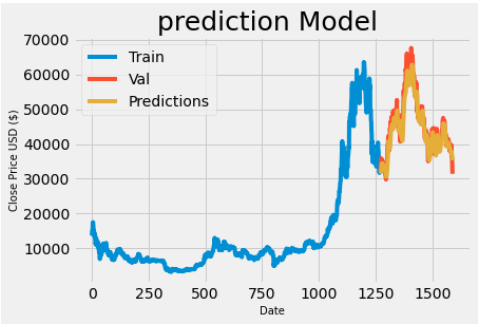
```
[array([0.16198395, 0.18257929, 0.18598206, 0.19217201, 0.22062372,
       0.22213934, 0.20582659, 0.18550172, 0.17656812, 0.1824425 ,
       0.15807598, 0.16701114, 0.17291197, 0.1637685 , 0.16451154,
       0.12830297, 0.12360997, 0.12806047, 0.13012016, 0.15020097,
       0.13000667, 0.11961186, 0.11863253, 0.12626504, 0.12471056,
       0.12334262, 0.12752884, 0.13290112, 0.12528572, 0.10678581,
       0.10857035, 0.09223958, 0.0869576 , 0.09230751, 0.0783498 ,
       0.05780358, 0.0702197 , 0.0681569 , 0.07817228, 0.08549997,
       0.08371107, 0.0760641 , 0.08844711, 0.08334435, 0.0972775 ,
       0.10772006, 0.10876934, 0.12243013, 0.11371103, 0.12418047,
       0.12695368, 0.11586555, 0.10521112, 0.10981394, 0.1022276 ,
       0.09992169, 0.11083369, 0.11641272, 0.11131869, 0.11991653,
       0.12202131, 0.12829053, 0.1286465 , 0.12959007, 0.11725681,
       0.10459819, 0.09572892, 0.09483572, 0.08750556, 0.09858327,
       0.09277712, 0.09261747, 0.07823788, 0.07872056, 0.07930332,
       0.07275165, 0.0775208 , 0.08384709, 0.08824347, 0.08848924,
       0.08536767, 0.08771728, 0.08442955, 0.08175055, 0.07729883,
       0.07144837, 0.07333613, 0.06107468, 0.05679705, 0.05808743,
       0.05607748, 0.05980156, 0.06558905, 0.05622687, 0.05556824,
       0.05284556, 0.05711681, 0.05886452, 0.05493494, 0.05593027])]
[0.058006439506896 1]
```

Building the LSTM Model

Training : We used Adam optimizer and Mean squared error for loss function.
Optimizer is used to improve upon the loss function, And the loss function is used to measure how the model did on the training.
Testing : Create an array contain a scaled data, y_test will be all the values that we want our model to predict so its actual test value, and x_test will be the past 100 value. We need to reshape the data because LSTM network expects the input to be 3-dimensional in the form of samples number of time steps and features which is the close price.

Predicting: we want predictions to contain these same values as our y_test dataset and we getting these predictions based of the x_test.

Evaluating: Evaluate the model by using mean root squared error (RMSE) its measure how accurate the model predicts the response and it's the standard deviation of the residuals and the lower value of RMSE is indicate a better fit. We get 1272 as a result, even though the model works perfectly great and the predictionswere exact the right values as they were supposed to be from the testing data .
Prediction model:



Looking at this graph we can see in blue line represents the data that the model was trained on and the reddish orange color it's the actual values for the rest of these days is the actual closing

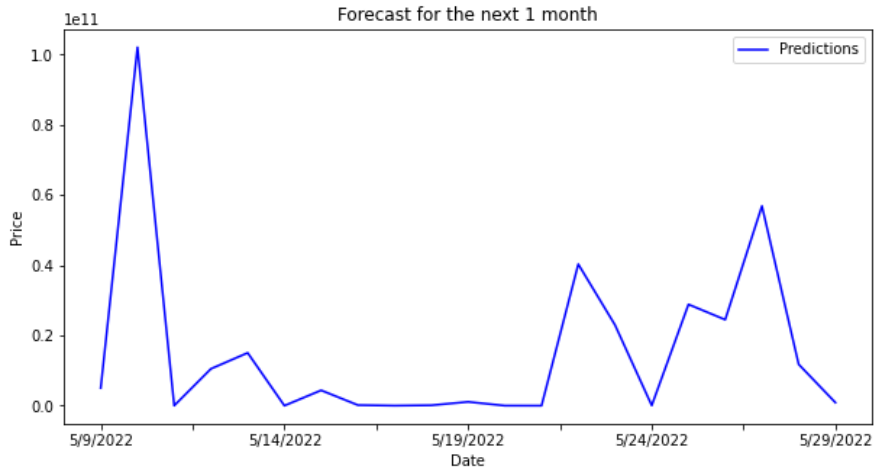
stock price for BTC for the rest of the days and then the orangish yellow is the predictions is what our model predicted the values to be, and we can see that they're really close so the model is pretty decent we can see that also in the table below:

Close	Predictions
32186.3	32220.336
34649.6	31959.54
34434.3	32330.164
35867.8	32870.645
35040.8	33707.586
33572.1	34329.473
33897	34412.402
34668.5	34304.957
35287.8	34283.88

3.**Random forest**

3.1Random forest Classifier

One of the most important questions among traders is if the Bitcoin currency will rise or fall, to determine the buying and selling period of traders to reduce risks in this case we choice Random Forest to predict the state of Bitcoin in the next month



The Fig shows the Bitcoin forecast for the next month, as we can see on 5/9/2022, the price will be high and the selling process will increase, and from 5/21/2022 to 5/21/2022, the currency will drop, and here we recommend buying Bitcoin

Buy price and date

Predictions	Date
19741464.42	2022 /14/5

Sell price and date

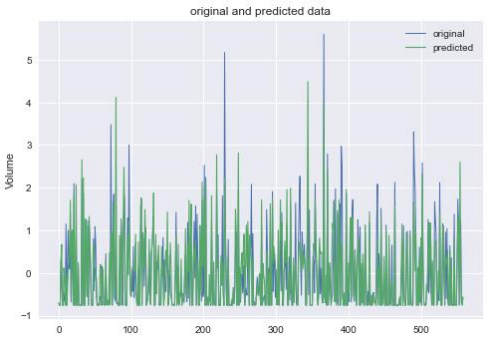
Predictions	Date
e+111.020580	2022/10/5

Date	Prediction price next month
5/9/2022	**5.05E+09**
5/14/2022	**19741464**
5/19/2022	**1.1E+09**
5/24/2022	**75404207**
5/29/2022	**903937153**

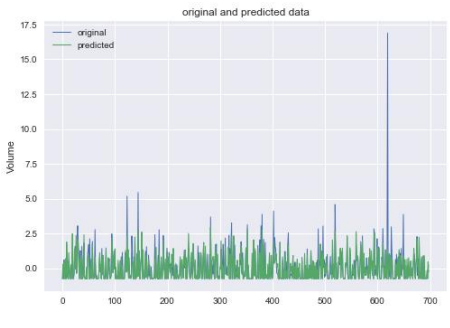
Predicting cryptocurrency is a challenging and interesting job for traders, investors and researchers because of the cost and complexity involved,The current study focuses on predicting the direction of the cryptocurrency 'Bitcoin', The predictive model proposed is built using ensemble learning via random forest. The model uses feature matrix drawn on technical indicators as input for learning and predicting. The proposed model is empirically evaluated and achieves good levels of accuracy Accuracy 98% which means that the ratio can be relied upon by traders.

3.2random forest regressor

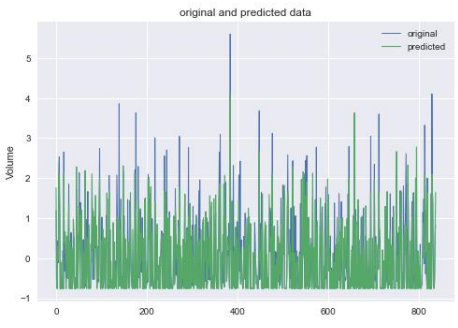
With the repetition of the model to learn a number of times according to the difference in the training sample and the test sample, the largest percentage obtained by the model is 96% with test_size=0.25



test_size=0.20



test_size=0.25



test_size=0.30

ethical and professional responsibilities

The data we used was high data available to everyone and does not contain any confidential or personal information, but we were keen not to tamper with the original data or falsify the results obtained from the algorithms.

Conclusion

Technical and quantitative analysts have applied statistical principles to the conventional financial market since its inception. Some attempts have been very successful, while others have been anything but. The key is to find a way to identify price trends without the fallibility and bias of the human mind specially in the cryptocurrency market .One approach that can be successful for investors is linear regression and random forest .We have downloaded the dataset and the important preprocessing has been done, after that we started to apply the models by using the historical prices of Bitcoin aiming to find an answer for our question which was: Can we predict the prices of cryptocurrency ?
The answer is clearly yes, and also based on the high accuracies we got, the model can be significantly useful for the investors who needs to take a decisions regarding to buy, sell, or hold

Both Multiple linear Regression and Random forest Classifier algorithms get a great result and high accuracy

Multiple linear Regression with 98.89%

Random forest Classifier with 98%