

Bit-coin price behaviour Prediction

* By Sayan Das



August 10, 2023

Contents

[Introduction 1](#_Toc142518807)

[Dataset 1](#_Toc142518808)

[Exploratory Data Analysis 1](#_Toc142518809)

[1) Plot of the closing price of bitcoin – 1](#_Toc142518810)

[2) Similar Values of closing bitcoin price and adjusted closing bitcoin price – 2](#_Toc142518811)

[3) The Distribution of Open, High, Low and Close Bitcoin Price – 2](#_Toc142518812)

[Feature Engineering 3](#_Toc142518814)

[1) Combining the High-Low and Open-Close 3](#_Toc142518815)

[2) Creating the target column 3](#_Toc142518816)

[3) Indicator for the quarter end 4](#_Toc142518817)

[Objective 4](#_Toc142518818)

[Model Training 4](#_Toc142518819)

[Conclusion 6](#_Toc142518820)

# 

# Introduction

Bitcoin, the world's first and most influential cryptocurrency, has witnessed tremendous growth and volatility since its inception in 2009. With its unique decentralized nature and the increasing acceptance in the mainstream financial ecosystem, Bitcoin presents a new frontier in financial market studies. Its highly volatile market behavior, along with the increasing amount of publicly available data, have made Bitcoin a particularly intriguing object of study for researchers and investors alike. Yet, the question remains: How predictable is Bitcoin's price?

This project is aimed to demystify Bitcoin's price fluctuations by adopting a straightforward yet effective analytical approach. While Bitcoin's behavior might seem intricate at first glance, our goal is to unravel its patterns using simple, comprehendible measures, and to construct a classification model that can predict whether an investment in Bitcoin would be profitable.

This project will draw upon historical Bitcoin data and employ well-established statistical techniques to study price patterns. By analyzing the key factors, we intend to gain a better understanding of Bitcoin's behavior. We will then leverage this understanding to build a predictive model. Rather than predicting the exact price, our model will classify whether Bitcoin would be a profitable investment during a particular timeframe.

# Dataset

The dataset used for the project, contains records for bitcoin prices for daily basis, for 2713 days from 17th September 2014 to 15th February 2022. This dataset has the following columns –

**Date** – The date for which the bitcoin price has been recorded (Dates, no missing observations)

**Open** – The opening price for the date (float, no missing observation)

**High** – The highest price for the day (float, no missing observation)

**Low** – The lowest price for the day (float, no missing observation)

**Close** – The closing price for the day (float, no missing observation)

**Adj Close** – The adjusted value of the closing price for the day (float, no missing observation)

**Volume** – The volume of bitcoin transaction for the particular day (integer, no missing observation)

# Exploratory Data Analysis

## Plot of the closing price of bitcoin –

From the time series plot of the closing bitcoin price plot it seems that the closing price of bitcoin has been increased over the years and the price reaches at its peak near the year 2021.

## Similar Values of closing bitcoin price and adjusted closing bitcoin price –

From the data it is observed that 100% of the values of the adjusted closing bitcoin price is same as closing bitcoin price of the corresponding day. Therefore, we conclude that the adjusted closing price have no additional information than closing bitcoin price.

Since the column has no additional information, it is not considered the analysis.

## The Distribution of Open, High, Low and Close Bitcoin Price –

From the histogram of the above-mentioned bitcoin price, it’s clear that all the bitcoin prices have a very highly positively skewed distribution that is right-tailed distribution.

# Feature Engineering

## Combining the High-Low and Open-Close

The columns High and Low are combined by taking the difference between Low and High and similar calculation has done for the columns Open and Close. After these two steps two new columns has been generated 1. low-high (Difference between Low and High price) and 2. open-close (Difference between Open and Close price)

## Creating the target column

A new column named as target has been created in the following way

The target value is an indicator variable that indicates whether in the next day Close bitcoin price will rise or fall, if target is 1 then it is profitable to invest on bitcoin.

From the pie chart of the target variable, one can see that around 45.85% cases the closing bitcoin price has gone higher in the next day.

## Indicator for the quarter end

A new column named as is\_quarter\_end has also been created that indicates whether the current date is a quarter end. This indicator variable takes value 1 for the months – March, June, September and December, otherwise it takes the value 0.

# Objective

In the project the objective is to find - if there is a way to predict whether tomorrows bitcoin price will rise or fall, so that one can invest on bitcoin. That is in technical, the objective is to find a model that can predict whether tomorrow’s bitcoin price will rise or fall.

# Model Training

To fulfill the objective a classification model is to fit taking target as response and the open-close, low-high, and the is\_quarter\_end as predictor.

Before predicting the model, the correlation among the predictors is checked. The correlation figures suggest that no significant correlation exists among the predictors that may create multicollinearity problems.

A logistic regression model has been fitted on the dataset taking target as response and low-high, open-close and is quarter-end as predictors. On fitting this model, the training accuracy is coming out to be 0.53 and validation accuracy is coming out to be 0.48.

# Conclusion

The fitted logistic regression model is giving a validation accuracy of around 48% for predicting the correct outcome. So, the model is no better than tossing an unbiased coin to decide whether to invest on bitcoin or not. The main, reason behind this poor performance of the classification model is the complexity of variables in the stock market. The price of bitcoin may be dependent on several other factors like, price of some other companies’ stock that are in a very important position as a competitor of bitcoin, may be the bitcoin price is dependent on other economic and stock market activities of the country, major economic and political event and so on. The data used for modelling, in this project are not sufficient in that sense. It can be reason behind the poor performance of the model.