Bitcoin Price Prediction

PRML Bonus Project Report

Abstract

This research is concerned with predicting the price of Bitcoin using machine learning. The goal is to ascertain with what accuracy the direction of Bitcoin price can be predicted.

Introduction

Bitcoin presents an interesting parallel to this as it is a time series prediction problem in a market still in its transient stage. As a result, there is high volatility in the market and this provides an opportunity in terms of prediction. In addition, Bitcoin is the leading cryptocurrency in the world with adoption growing consistently over time. Due to the open nature of Bitcoin it also poses another paradigm as opposed to traditional financial markets.

Dataset Description

The goal of this is to ascertain with what accuracy the future price of bitcoin can be predicted. The dataset provides opening, closing, highest price, lowest price, volume and market cap of a particular day.

Attributes

1. “Open” : Opening price of bitcoin on a particular day.
2. “Close” : Closing price of bitcoin on a particular day.
3. “High” : Highest price of bitcoin on a particular day.
4. “Low” : Lowest price of bitcoin on a particular day.
5. “Volume” : Volume of bitcoin in the market on a particular day.

6. “Market Cap” : Market capitalization price of bitcoin on a particular day.

Approach

We start off by importing the dataset and extracting all the useful columns from the dataset. Now, we add a feature, which is to be predicted, containing the closing price of bitcoin after 10 days. We add one more feature containing the average value of all the 4 selected features, so that we have a better approximation. We preprocessed our data over here and trained the final database after splitting the dataset into training and testing with a ratio of 85:15. After the transformation of our dataset, we will be using different regression models to find the r2_score of the data for each of them.

Building Models

We identified the best hyperparameters for each model through RandomizedSearchCV before implementing that model, so that we get a better result.

1. Linear Regression model

   r2_score : 0.9595479455040584

   Hyperparameters : None

2. KNeighborsRegressor

   r2_score : 0.982804030893062

   Hyperparameters : (n_neighbors=12, weights = 'distance', p=2, leaf_size = 36)

3. RandomForestRegressor

   r2_score : 0.981511118417228

   Hyperparameters : (n_estimators = 227, random_state = 0,

max_features = 0.3, max_depth = 15, criterion = 'absolute_error', min_samples_split = 2, min_samples_leaf = 3)

4. DecisionTreeRegressor

   r2_score : 0.980426628891348

   Hyperparameters : (random_state = 0, max_features = ‘log2’, max_depth = 14, criterion = 'friedmen_mse', splitter = ‘random’, min_samples_split = 5, min_samples_leaf = 2)

5. XGBRegressor

   r2_score : 0.9827432612848608

   Hyperparameters : (n_estimators = 40, max_depth = 6, learning_rate = 0.25)

6. LGBMRegressor

   r2_score : 0.9818203184051063

   Hyperparameters : (num_leaves = 120, max_depth = 5, n_estimators = 120, learning_rate = 0.13, cosample_bytree = 0.7, random_state = 0)

Building Pipeline

After implementing all the models and analyzing every model, I selected best 5 models to be implemented in a pipeline, namely,

- KNeighborsRegressor
- RandomForestRegressor
- DecisionTreeRegressor
- XBGRegressor
- LGBMRegressor

Results

After implementing all these in a pipeline, I got a final r2_score of 0.9838108554888297.

Extra Efforts

I predicted the prices of bitcoin for 10 days which are not in the dataset, and appended the value in the feature, which was to be predicted and then visualized the predictions and the dataset values through a graph with respect to the change in dates.

Moreover, I deployed the .pkl file, which I got through the pipeline, within the colab file. That cell gives the prediction of bitcoin price after 10 days from the date of given data.