Pstat231 Project Memo

Zihao Yang

2022-04-10

Dataset

Recently I'm interested into the Cryptocurrencies. So I choose the dataset called Cryptocurrency Historical Prices from Kaggle (link here). As its name shown, it is a dataset that includes the historical prices information of some of the top Cryptocurrencies by market capitalization. It contains one csv file for each kind of Cryptocurrencies. There are 23 files of different currencies and each of them containing hundreds to thousands of observations. Each of the observation is the price history on a daily basis, starting from April 28, 2013.
There are 7 predictors:
1.Date : date of observation
2.Open : Opening price on the given day
3.High : Highest price on the given day
4.Low : Lowest price on the given day
5.Close : Closing price on the given day
6.Volume : Volume of transactions on the given day
7.Market Cap : Market capitalization in USD
Although the data of the price history should be a time series or continuous data, it has been reorganized by the daily basis, so it should be discrete data which will be easy for further operations. Since different currencies showed up in different time, some new Cryptocurrencies have less observations than old ones because they appearences are shorter than others. It may cause missing values, if I choose a relatively long time period. To address this problem, I need to make sure that all the Cryptocurrencies should have earlier appearance time than the time period I chose, when I try to work on the correlations of some currencies.

Research questions

I'm interested in predicting the future price of a certain Cryptocurrency. Also, I want to know:
1.How did the historical prices of various currencies change over time.
2.Which kinds of Cryptocurrencies are more stable or volatile?
3.Which can make large profit during a short period? What about a long period.
4.How does the price fluctuations of currencies correlate with each other?
5.Does there exist any pattern with weekly, monthly, seasonally, and yearly fluctuations?
I think the regression approach would be a good start on figuring out these questions. All of the predictors will be helpful, because they are close to the price change and commonly used for predicting prices. The goal of my model should be a combination, because first I want to check the pattern of the historical prices and then I want to predict the future price. Thus, it should be a descriptive and predictive model.

Timeline

A general timeline will be:
Week3-Week4: Clean the data and construct valid models based on the price history.
Week5-Week6: Determine the best fitted model, find if there is any pattern, and analyze the data.
Week7-Week8: Train the model, predict the future value, and find the correlation of different currencies.
Week9-Week10: Refine the model, improve the accuracy, continue finding the correlation of different currencies.
Week11: Wrap up everything, make the conclusion, and answer the research questions.

Concerns

So far I don't have any questions or concerns, but if I have some when working on the project, I'll go to the office hour or ask on Piazza immediately.