I did a thing,

So I was bored over the long weekend and decided to play with data a little bit, so I am doing some light reading and I come across this article about time series forecasting.

<https://towardsdatascience.com/time-series-modeling-using-scikit-pandas-and-numpy-682e3b8db8d1ut>

It is so interesting I could not just read and pass I decided to play around with something similar but with cryptocurrency data, so run to Kaggle to look for a dataset and right as rain I find on that contains historical prices from 2013 to July 2021 here

<https://www.kaggle.com/datasets/sudalairajkumar/cryptocurrencypricehistory>

It contained 9 columns:

SNo, Name, Symbol, High, Low, Open, Close, Marketcap

Looks something like this (I dropped some columns):

Text, table

Description automatically generated

To begin I started some feature engineering, very simply, I just added extra columns that captured the values for the previous day like so:

Text, application

Description automatically generated

Then I split the data into training and testing. I used all the data from 2013 to 2020 for training and then used the 2021 data for testing.

Now to the good stuff,

Before the actual and final prediction, trained 3 different models to see which one would perform the best

Linear Regression,

Random Forest and

K-Nearest Neighbor

But, and this came as a surprise to me, the Linear Regression performed way better than the other 2 as seen below

Chart, box and whisker chart

Description automatically generated

Nearly perfect (suspicious you would say yh?)….lol

So it is decided, I shall use Linear Regression but not without some Hyperparameter tuning so I pick up my phone and call my good friend GridSearchCV.

I run a few iterations with some parameters and an 8-Fold Cross-Validation algorithm using the same custom scoring function in the article from before and arrive at the best model

Graphical user interface, text, application, email

Description automatically generated

Muahahahahahahahaha soon my master plan will be complete and I shall take over the world…..

Just kidding….lol

Now lets test and see what we have, I predict using the test part of the dataset (2021 Prices) and what do you know, I have an almost perfect match

Chart, line chart

Description automatically generated

Now I know what you are thinking….This is too perfect, a model cannot be that accurate. I thought so too so I decided to find another dataset and use it to test. I found one here

<https://ng.investing.com/crypto/bitcoin/historical-data>

So I run my feature engineering on it again to prepare it for prediction (Adding columns for the previous day figures)

Now I pass the new data set into my Perfect Model aaaaaaaannnddddd

Drum rollllllllllllllllllllllll

I still get a pretty impressive prediction although not as accurate as the first test with much higher error rate.

Chart, line chart

Description automatically generated