BTC Time Series Analysis with Machine Learning

img: https://www.roinvesting.com/en/crypto-articles/how-to-trade-bitcoin

Bitcoin is a digital currency not published by a central bank. Recently, in some countires, they are trying to accept Bitcoin as their economic currency as well. Also, bank interest rates are too low to invest capital. People started to turn away their eyes on investing their money to Stock market or Bitcoin market. There are many factors affecting market price, but time series attempt to predict the future price based on the past prices. From this report, I will try to predict whether coin price will go up or down based on the past price information with one of the machine learning algorithms called XGBoost Classification.

1. Data Explanation

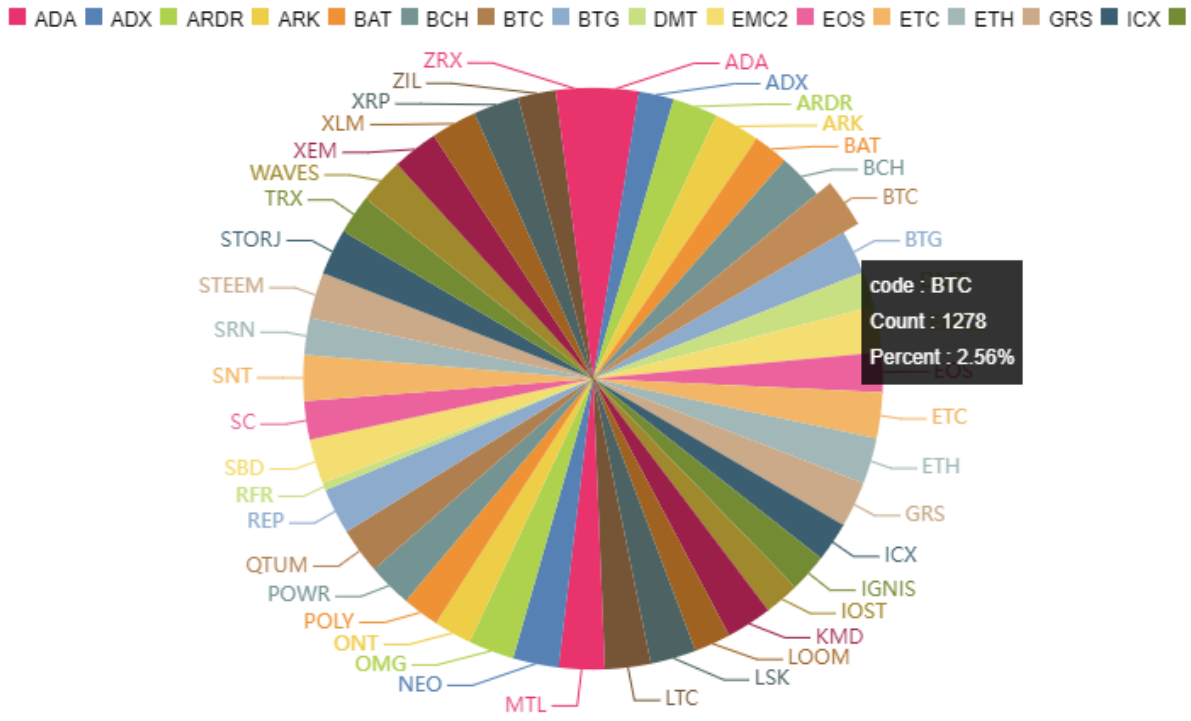
	code	candleDateTimeKst	highPrice	lowPrice	tradePrice
1	BTC	2021-03-25T09:00:00	65305000	62727000	63174000
2	BTC	2021-03-24T09:00:00	68370000	64500000	64777000
3	BTC	2021-03-23T09:00:00	66279000	63000000	65458000
4	BTC	2021-03-22T09:00:00	68380000	64345000	64725000
5	BTC	2021-03-21T09:00:00	68715000	65451000	67120000
6	BTC	2021-03-20T09:00:00	69980000	67700000	67849000

Data Table Head

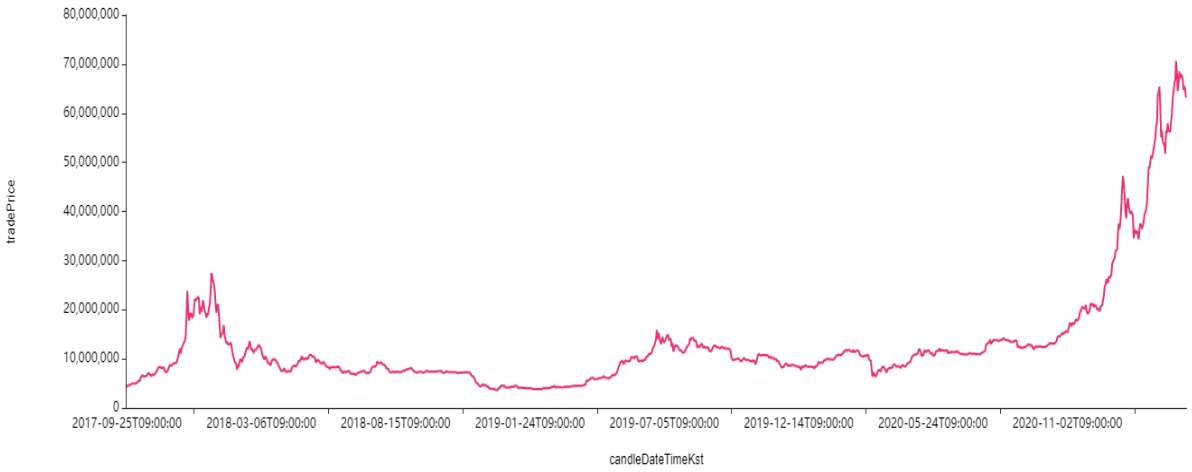
Variable Explanation

Variable	Exaplanation
code	Name of coin ex) BTC = BitCoin
candleDateTimeKst	Date and Time in Korea
highPrice	The hightest Price of each coin on certain day
lowPrice	The lowest Price of each coin on certain day
tradePrice	The last Price of each coin on certain day

Types of Coins in Dataset



There are plenty of different types of coins in this dataset. Overall 114 types of coin are listed in the data set. Before we started to analyze our data, every coin has different duration. For example, 'BTC' coin which stands for BitCoin started to record from 2017-05-25 to 2021-03-25. Hence, 'DOGE' coin started to recored from 2021-02-24. In the case of DOGE coin, only a month data set exists in this data set. When we divide data set into train and test data sets. Some of coins have not enough data to train. Therefore, we will remove some of coins which have amount of records less than a year. Finally, we have 77 different coins have recoreded more than a year.



BTC Trade Price changes between 2017-09-25 and 2020-03-25

Nextly, we will divide data set into train and test set. All of 77 coins have different starting points, but have the same end points on 2021-03-25. Since we already removed some of coins having their starting point after 2020-01-01, we will be able to train at least a year of records if we divide train and test data set before 2021-01-01. Conclusively, we will use our training data set before 2021-01-01 and our test data set after 2021-01-01.

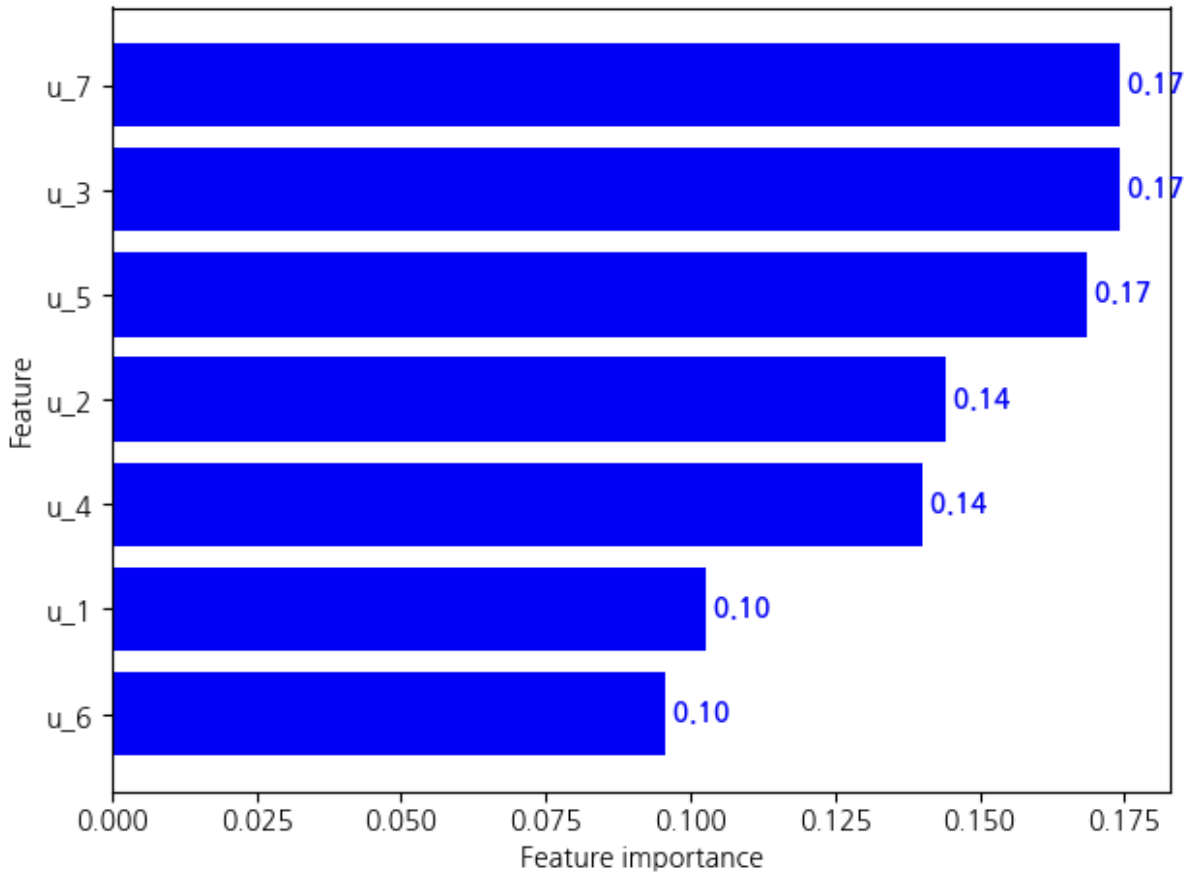
Mostly, supervised machine learning problems classfied into two problems. One is 'Classification' and the other one is 'Regression'. We can't generally say which one is eaiser than the other, however, mostly classification problems make problem eaiser than when we consider them as regression. It is much easier when we attempt to predict whether coin price has gone up or down on the day compared to yesterday, than just trying to predict the exact price difference. Therefore, our goal in this case is to predict whether coin price increases or decreases than yesterday based on the recoreds during 7 days.

XGBoost Classifier is one of the machine learning algorithms showing great performance for most problems. The explanatory variables are the records whether trade price increases or decreases on certain day.

u_1	u_2	u_3	u_4	u_5	u_6	u_7
1	1	1	1	1	1	1
0	1	1	1	1	1	1

Explanatory variables

Our new columns from u_1 to u_7 represents whether trade price increases or decreases. For example, if u_1 is 0, then it means that today's trade price decreased from yesterday's price and if u_2 is 1, then it means that today's trade price increased from two days before. In the same manner, target variable also refactored into 0 and 1. It is 0 if the tomorrow's price decreases from today's price and vice versa.



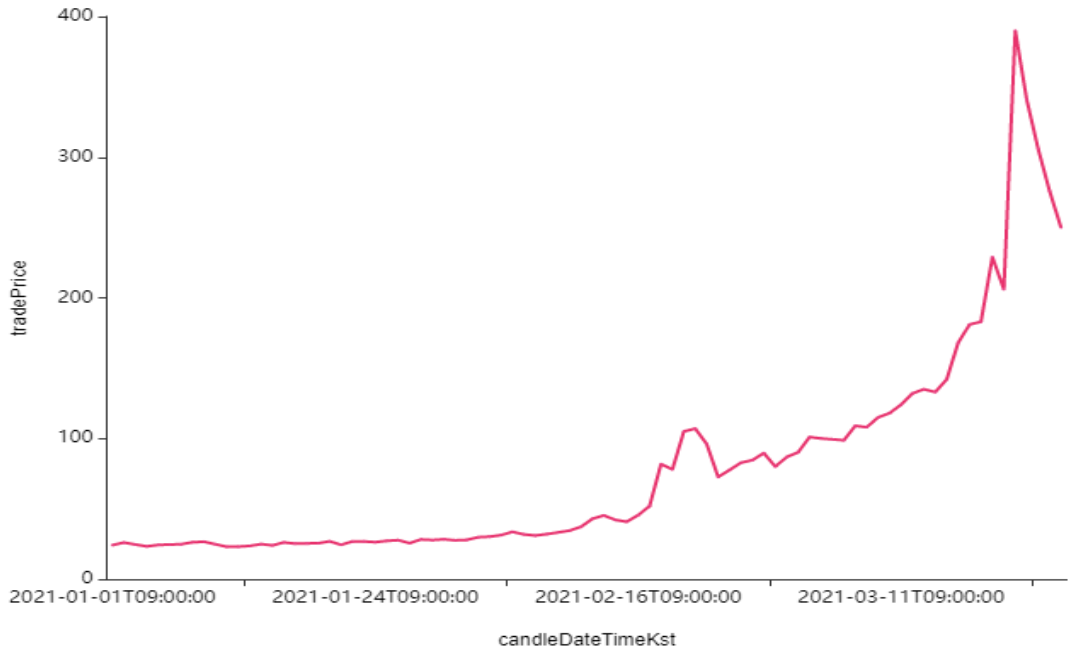
BTC coin Feature Importance

The plots above shows the feature importance of XGB Classification on BTC coin. It tells us that prices on 3 and 7 days ago are the most important feature to predict tomorrow's coin price. There are 77 different XGB classification models on 77 different coins. Mostly, price on three days ago is the most important feature among 77 coins.

	code	label	f1	precision ↓	recall
1	UPP	1	0.410958...	0.789473684...	0.27777777777...
2	DMT	1	0.55	0.785714285...	0.42307692307...
3	HBAR	1	0.530120...	0.758620689...	0.40740740740...
4	ONT	1	0.531645...	0.75	0.41176470588...
5	MANA	1	0.48	0.75	0.35294117647...
6	TSHP	1	0.363636...	0.75	0.24
7	KMD	1	0.447368...	0.739130434...	0.32075471698...
8	IQ	1	0.556962...	0.733333333...	0.44897959183...
9	POWR	1	0.561797...	0.714285714...	0.46296296296...
10	ADX	1	0.536585...	0.709677419...	0.43137254901...

Top 10 Precision Coin list

The table above tells us top 10 coin list having higher precision. UPP coin has precision approximatly 0.79. It looks like high enough to invest all assets! However, there is Some tricks. As line plot below, UPP coin mostly increases during the test period. Since it is already imbalanced data, we will not able to say that our prediction was precise and accurate.



In conclusion, time series problem applies many real world problem like stock market and coin market and so on. At the same time, real world problem has various factors affecting our predictions and inferences. In this case, dealing with imbalanced data and situation with lack of data is one of the main issues.