

Cryptocurrency Price Speculation & Prediction

Thip Rattanaivilay
DSC680 - Fall 2021 – Project 2
<https://github.com/thiprattanaivilay/DSC680>

Which Domain?

During Covid lock down, I have spent the entire year of 2020 understanding Blockchain and Cryptocurrency, especially Bitcoin and Ethereum. These two crypto are the most volatile markets today and has gained a lot of attention from investors across the globe. Cryptocurrency, being a novel technique for transaction system, has led to a lot of confusion among the investors and any rumors or news on social media has been claimed to significantly affect the prices of cryptocurrencies. The goal of this study is to predict prices for crypto coins using Machine Learning Techniques for the few I weeks for project 2 I will prepare a strategy to maximize gains for investors, friends and of course myself. I also aim to find out if there is a co-relation between fluctuating crypto prices and related news. Here I have list 10 references that I will be gathering my information from to help with this project.

1. *Cryptocurrency prices, charts and market capitalizations*. CoinMarketCap. (n.d.). Retrieved October 3, 2021, from <https://coinmarketcap.com/>.
2. Polygon.io. (n.d.). Retrieved October 3, 2021, from https://polygon.io/crypto?gclid=CjwKCAjwhuCKBhADEiwA1HegOT3u7UXGs_N_zCyrQ06X8blMS1oLFdzB4pILdbwyFwI6XzqPnJBxoCrygQAvD_BI.
3. *Crypto Exchange*. Gemini. (n.d.). Retrieved October 3, 2021, from https://www.gemini.com/exchange?utm_source=google&utm_campaign=USA_Search_CryptoGeneral_SKAGS_New_June&utm_medium=cpc&utm_content=126011091779&utm_term=cryptocurrency&gclid=CjwKCAjwhuCKBhADEiwA1HegOczjgj8rNgOMg72tqZlcnD5vkmjkHvBoAz2RaxP1Ze9gNFSZq5jylhoCKgYQAvD_BI.
4. Sigalos, M. K. (2021, October 2). *Ethereum had a rough September. here's why and how it's being fixed*. CNBC. Retrieved October 3, 2021, from <https://www.cnbc.com/2021/10/02/ethereum-had-a-rough-september-heres-why-and-how-it-gets-fixed.html>.

5. Yahoo! (n.d.). *The Crypto Daily – Movers and Shakers – october 2nd, 2021*. Yahoo! Finance. Retrieved October 3, 2021, from <https://finance.yahoo.com/news/crypto-daily-movers-shakers-october-002300815.html>.
6. Srk. (2021, July 7). *Cryptocurrency historical prices*. Kaggle. Retrieved October 3, 2021, from <https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory>.
7. Taniaj. (2018, June 3). *Cryptocurrency price forecasting*. Kaggle. Retrieved October 3, 2021, from <https://www.kaggle.com/taniaj/cryptocurrency-price-forecasting>.
8. Kash. (2021, September 11). *Ethereum Cryptocurrency Historical Dataset*. Kaggle. Retrieved October 3, 2021, from <https://www.kaggle.com/kaushiksuresh147/ethereum-cryptocurrency-historical-dataset>.
9. Sudalairajkumar. (2017, November 8). *Cryptocurrency data pull*. Kaggle. Retrieved October 3, 2021, from <https://www.kaggle.com/sudalairajkumar/cryptocurrency-data-pull>.
10. *CoinDesk: Bitcoin, Ethereum, crypto news and Price Data*. CoinDesk Latest Headlines RSS. (n.d.). Retrieved October 3, 2021, from <https://www.coindesk.com/learn/>.

Which Data?

For the purpose of this research, I will utilize data from Jan-2018 to Aug-2019, concerning the hourly prices in USD and were divided into training set consisting of data from Jan-2018 to Feb-2019 (10176 values) and testing set from Mar-2019 to Aug-2019 (4416 values). This data was taken from Kaggle.com, kraken.com website and others like Coinbase etc., which is a trading platform and data repository for cryptocurrency exchanges. Also, I will utilize those data for forecasting and predictions (number of past prices taken into consideration).

Variables:

Open

Close

Volume

Volume Currency

Weighted Price

Date

Crypto Variables:

BTC

ETH

LTC

XRP

ETC

STR

DASH

Dataset link:

Srk. (2021, July 7). *Cryptocurrency historical prices*. Kaggle. Retrieved October 3, 2021, from <https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory>.

Taniaj. (2018, June 3). *Cryptocurrency price forecasting*. Kaggle. Retrieved October 3, 2021, from <https://www.kaggle.com/taniaj/cryptocurrency-price-forecasting>.

Kash. (2021, September 11). *Ethereum Cryptocurrency Historical Dataset*. Kaggle. Retrieved October 3, 2021, from <https://www.kaggle.com/kaushiksuresh147/ethereum-cryptocurrency-historical-dataset>.

Sudalairajkumar. (2017, November 8). *Cryptocurrency data pull*. Kaggle. Retrieved October 3, 2021, from <https://www.kaggle.com/sudalairajkumar/cryptocurrency-data->

Research Questions? Benefits? Why analyzes these data?

Cryptocurrencies, to be specific, has gained a lot of traction in the recent years from investors accross the globe. Cryptocurrency being a novel technique for transaction system has led to a lot of confusion among the investors and any rumors or news on social media has been claimed to significantly affect the prices of cryptocurrencies. Bitcoin is one of the oldest and biggest cryptocurrencies being traded as of now, in terms of the volume being traded. It is so big that even now, with the advent of thousands of new cryptocurrencies, Bitcoin has a market share of more than 55% as compared to other cryptocurrencies, being followed by Ethereum at 8.57%. This says a lot about why Bitcoin might be an interesting and important stock to predict. Also, Bitcoin prices fluctuate heavily. Over the past 2 years, Bitcoin has seen its highest price around \$47,353.14 and its lowest price around \$900. It is very sporadic, and this is one of the most important reasons which attracted us to analyze and predict its price.

Research Question:

How do cryptocurrency markets look like int he coming days, weeks or years?

What are the causes dips and spikes in crypto values?

Are the markets for different altcoins or largely independent affect the market cap for crypto?

How can I predict what will happen next?

What Method?

To solve the problem using Machine learning, I will first try to categorize the problem and tried to find previous solutions on how they solved it. I quickly learned that, since the problem involves prices which are changing with time, this could be modelled as a Series prediction problem. In parallel, I also tried to solve the problem as a normal machine learning problem with the features being the previous prices and the output being the price predicted for that day. I explain what models I have used and how I configure them to predict the crypto prices.

CRIDP DM Process:

- Business Understanding.
- Data Understanding.
- Data Preparation.
- Modeling.
- Evaluation.
- Deployment

What model will I use for this project?

ARIMA: Autoregressive integrated moving average (ARIMA) is a statistical regression model, which can be utilized in time series forecasting applications, such as finance. ARIMA makes predictions while considering the lagged values of a time series, while accommodating for non-stationarity. The model, which is one of the most popular linear models for time series forecasting, originates from the autoregressive (AR) and moving average (MA) models, as well as their combination, also known as ARMA (Wikipedia).

Random Forest: Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees (Wikipedia).

Recurrent Neural Net: RNNs provide a generalization of the feed forward network model for dealing with sequential data, with the addition of an ongoing internal state h_t serving as memory buffer for processing sequences (Wikipedia)

Potential Issues?

There are many obvious issue that comes to mind is how to deal with wildly different results from using different approaches that might have many limitations. I do fear that I won't complete the coding portion effectively with the given timeline. The data found are both numerous and high quality. There's no missing or fails data that needs to be removed, and some of the noise where a wait time will take longer and during a couple minutes is smoothed by the prediction model.

Concluding Remarks

The goal of this project 2 is to provide an easy introduction to cryptocurrency analysis using Data Science. I will walk through a simple data mining and cleaning which is retrieve, analyze, and visualize data on different cryptocurrencies. In the process, I will uncover an interesting

trend in how these volatile markets behave, and how they are evolving and how I can predict the outcome of crypto.

Reference:

https://en.wikipedia.org/wiki/Random_forest

https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average

https://en.wikipedia.org/wiki/Recurrent_neural_network