# Machine learning NanoDegree

# Capstone Proposal

Name: kelvin

E-mail: Kelvinwj98@gmail.com

17/09/2021

## Domain background

Nowdays cryptocurrency has become one of hottest topic ,it rivals  traditional currency since cryptocurrency has advantage as being decentralized,in a decentralized world, game-theoritical incentives take over the role of a general, governing authority.

And now a lot of people has trying to invest in Cryptocurrency (Bitcoin,Ethereum and BNB),but the problem is commonly people don’t know how to valuing a coin,according to www.forbes.com/ 90% trader lose their money since it takes a lot of effort to mastering how to invest and trade.The goal of this project is to attempt to predict the price of coin  in the next few days using machine learning algorithm to help people that trying to making money from cryptocurrency.

## Problem Statement

---

The problem is we need to make the model to predict the price for a given information in the about the price behavior in the past. I will classify this as timeseries problem,and since the predicted output is price(continuous value) ,we will use regression technique to handle it .

## Datasets and inputs

---

The datasets are provided by CoinMarketCap on Kaggle competition website.the datasets contains 2991 rows and 10 columns.

Input data fields:

- SNo- The id of training set.(wont be used)
- Name- The name of coin.(wont be used)
- Symbol- The abbreviation that contains 3 character from the coin (ex: Bitcoin).(wont be used)

  Three features from above will be deleted, both name and symbol will be deleted since its timeseries data , all name and symbol are all the same .and Sno also wont be used since dataframe already provide identifier.

- Date- A trade date.
- High- The highest price at which a stock traded during a period.
- Low- The lowest price of the coin on that date.
- Open- Starting period of trading on a securities exchange.
- Close- End period of trading on a securities exchange.

- Volume- Volume is the amount of an asset or security that changes hands over some period of time, often over the course of a day.Marketcap- total dollar market value of a coin's.

Example:

| Sno | Nme | Symbol | Date | High | Low | Open | Close | Vol |
|---|---|---|---|---|---|---|---|---|
| 1 | Bitcoin | BTC | 21/9/201 | 9000 | 8450 | 8500 | 8600 | 1000 |
| 2 | Bitcoin | BTC | 22/9/2011 | 8800 | 8100 | 8240 | 8405 | 1200 |

## Solution statement

---

The solution we provide for this problem is to apply RNN-LSTM for Multivariate Timeseries forecasting.

First, I will try to process the dataset and get the visualization of the data to get some insight.

Then I will try to extract features from datasets and delete useless features such as Coin Name,Symbol and No.

## Benchmark model

For this problem, the benchmark model will be using RNN-LSTM , and compare it with other multivariate timeseries forecasting algorithms.

## Evaluation metrics

Prediction results are evaluated on Threshold metric(Accuracy)

## Project design

Before,I do anything to the data, I will first check the dataset features ,dataset shape and how each features are formatted. Then, I will do data cleaning to erase a row if any of the column contain Nan value.

In this dataset, since its only has 4 input data we don't need to use any dimension reduction technique and I wont use encoding because all the input data are numerical data.

And next , I will visualize the data with line plot to get better understanding of the data distribution.

Then next, I will train the data ,first method would be using ARIMA and then next method would be RNN-LSTM, then I would compare the result which method give higher accuracy is the final model to be submitted.

Reference:

1,https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory

2, https://www.investopedia.com/terms/c/cryptocurrency.asp