

# Capstone Project

## Forecasting cryptocurrency price

### Domain background

A cryptocurrency is a digital asset designed to work as a medium of exchange wherein individual coin ownership records are stored in a ledger existing in a form of computerized database using strong cryptography to secure transaction records, to control the creation of additional coins, and to verify the transfer of coin ownership. Cryptocurrencies typically use decentralized control as opposed to centralized digital currency and central banking systems (Wikipedia).

Recently, I have been in touch with this field and I have seen an opportunity to apply time series forecasting to the data.

Other work on this topic has used LSTM models [1][2][4][5], LightGBM [3] or XGBoost [4].

### Problem statement

Cryptocurrency prices (how much USD dollars is one unit) are changing every minute and are usually very volatile. People invest in cryptocurrencies by buying and selling afterwards when they have benefits. The problem is knowing what will be the price of the cryptocurrency in the future. Time series forecasting models could be used to make predictions based on past data to help make decisions on how to invest.

### Solution statement

As a solution, a forecasting machine learning algorithm will be trained to be able to predict the price of cryptocurrencies in the near future based on past data. Predictions can be later compared to real values.

The model proposed for predicting the future price of a cryptocurrency is to use the SageMaker [DeepAR Forecasting Algorithm](#).

### Datasets and inputs

For the dataset, [historical cryptocurrency price data](#) will be used. The dataset contains the daily data for 23 different cryptocurrencies. For the case of bitcoins, there is data from 2013-04-29 to 2021-02-27.

Columns:

- SNo: index
- Name: cryptocurrency name (e.g Bitcoin)

- Symbol: cryptocurrency symbol (e.g BTC)
- Date: timestamp, each row has the values for one day
- High: highest price achieved of the day
- Low: lowest price achieved of the day
- Open: opening price of the day
- Close: closing price of the day
- Volume: volume of transactions on the given day
- Marketcap : Market capitalization in USD

The target time series will be the closing price of the day.

## Benchmark model

The benchmark model will be a linear learner (also available in AWS Sage Maker).

## Evaluation metrics

As evaluation, the Root Mean Squared Error (RMSE) will be used to compare predictions with actual values.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Predicted_i - Actual_i)^2}{N}}$$

The DeepAR algorithm also gives a confidence interval for the prediction.

## Project design

The steps of the project are:

1. Downloading the dataset
2. Analysing the dataset and data visualization
3. Dataset processing
  - a. Define context and prediction windows
  - b. Define time series
  - c. Requirements: Stationarize the series: detrending, differencing and seasonality
4. Creating train, validation and test datasets
5. Training and optimizing the model
6. Testing and evaluation of the model
7. Deploy the model
8. Create a simple app to display predictions done with the deployed model
9. Create a final report

First the linear model will be trained and evaluated to have a baseline. After that, the DeepAR model will be trained and optimized using hyper-parameter tuning to try to improve the results.

Also, different models will be trained for each (or some) of the cryptocurrencies to see how they perform for each one. Using the DeepAR categorical features, one single model could be trained with all the data from all the cryptocurrencies.

## Sources

- [Wikipedia - Cryptocurrency](#)
- [What Determines the Price of 1 Bitcoin?](#)
- [Kaggle](#) dataset
- [Cryptocurrency Price Prediction Using Deep Learning](#)
- Udacity time series forecasting [example](#)
- AWS SageMaker [DeepAR Forecasting Algorithm](#)
- AWS SageMaker time series forecasting with linear learner [example](#)
- AWS SageMaker DeepAR [example](#), [example](#), [example](#)