Given the cryptocurrencies’ data, we aim to forecast the future cryptocurrencies’ prices so as to execute profitable trades. We show that the cryptocurrencies’ prices also exhibit desirable properties such as stationarity and mixing. Some classical time series prediction models that exploit this behavior, such as “Arima” models produce poor predictions and also lack good probabilistic interpretations. We have introduced a theoretical framework in the 1st place and for predicting and trading prices of the cryptocurrencies for future and based on that we have designed our model which is based on “Neural Network” model which can give better prediction values as compared to the other models.

**Cryptocurrency Case**

**KAUTILYA Team Mentors:**

* Subhabaha Pal(@drsubhabahapal)

**KAUTILYA Team:**

* Prakash Kumar (@prakash), MAHE, India
* Sujit Mohapatra(@sujit), MAHE, India
* Vivek Arya (@svarya), MAHE, India
* Amit Kumar Mohapatra (@amit98), MAHE, India
* Vijay (@vijays), MAHE, India

**Team Toolset:**

* R-Studio: Neural Network Time Series Forecasting
* MS Excel

**Business Understanding**

* Who: Data Science Society (Academic Datathon) Organizers
* What (1): Make a prediction model of the major cryptocurrencies’ prices
* How: Intra-group brainstorming based on various concepts of time series analysis and machine learning.

**Data Understanding**

* Explore dataset’s structure
* Discover potential variable dependencies across datasets
* Identify the missing data points across the dataset
* Identify the subsets of data that modeling would be based on

**Data Preparation**

The cryptocurrencies data had many NA values and deleted (absent) rows for particular time points, which are common in any financial time series. So we created time points with NAs in rows using Microsoft Excel. The NAs were then imputed using a function from” imputeTS” library of R. The total dataset was then subset to create the desired model.

* For the target forecast, two approaches opted: the forecast to be made on per 5 minutes data of each day as the cryptocurrencies market works on 24\*7. The two approaches on which we have worked on are i. “auto.arima”, ii. “nnetar”.We ensured that there are no variables with a large percent of missing/outlier values.
* The prices of cryptocurrencies for 25th Jan 2018 with 288-time points were predicted.

**Modeling**

The subset of the dataset had price values of currencies for the month of January. The starting time point was taken as 18th Jan 2018 00:00 and end time point was taken to be 24th Jan 2018 23:55 for training the model and which predicts the 25th Jan 00:00. Keeping these things in mind the model predicts the currencies’ prices one by one.  The model has been formed as per the following procedure and so on.

https://www.datasciencesociety.net/wp-content/uploads/2018/04/Picture1-3-300x50.png

**Problem Faced**

* The main problem was with the dataset which is not continuous due to which we have to hit and trial some methods for missing data treatment and imputation.
* Due to the missing values for which the data points have been imputed, the actual values may vary as in some cases the data points are missing for 2 or more time points.