

Project 1: Bitcoin Price Prediction

Due Date: *Monday, September 29 (11.59PM)*

Submission: on Gradescope

Instructions:

The projects may be completed in groups (up to 4 students). If you do so, please make sure that you include everyone's full name, and that you *also select everyone's name when submitting the assignment on Gradescope*. This ensures that each group member will get a grade assigned and have access to the comments from the graders.

For this project, the following items need to be submitted:

1. the **code** in R or *Python* (a Markdown file is ok)
2. a **report** (as a pdf file) that summarizes the process you followed for reconstructing and analyzing the data. The report **should not contain** any code, but it should have a snapshot of the reconstructed data, appropriate summary plots, and other metrics used in your analysis and decision making process. The report should include: (i) an introduction section, (ii) the data cleaning process, (iii) the methodology/approach you took to analyze the data and perform model selection and (iv) all appropriate results. Finally, the report should end with a conclusion where you summarize your findings. The report should be around 5 pages long.

Project Description:

The project is inspired from a Kaggle competition [here](#) and the data can be downloaded from Coursera. Please read the Kaggle description to familiarize yourselves with the data set. The main **goal** of this project is to **predict the bitcoin market price**, `btc_market_price` using techniques that have been discussed in class up to now.

A few things to consider are the following:

1. You will need to split the data into training and testing. Be careful because the data are longitudinal, so you should not do a random split.
2. Because of the nature of the data, you will also need to use information from the previous day/days to predict the market price at a future date. For example, on day 1, you will use the information from days 1,2,3,4 etc. to predict the market price on day 7. I will

leave it up to you to decide how many days to use for the prediction task. Keep in mind that the market price is not the only one that has a longitudinal structure, so be mindful when deciding the cut-off. (If your algorithms do not run, consider using 2 or 3 days.)

3. You will probably need to reconstruct the data.
4. The data is not perfect, so you will also face challenges such as missing data, categorical predictors, outliers, scaling issues, computational issues, and others. Use your best judgment to deal with these. **There is not on best answer**, but you need to *justify all the decisions that you make*: this is what we will be looking for when grading.

In the end, you need to summarize your process and results in a report that does not contain any code.

Grading Rubric:

Data Configuration: 30 points

Model Selection and Prediction: 30 points

Report Structure and presentation: 20 points

Code Structure: 20 points