

"TEMPUS

Data Science Analyst
Interview Case 1
July 2016

Background:

At Tempus we are constantly evaluating the data we have access to, to see how it can be clinically informative. We would like to see an example of your statistical process of the attached data for your examples of your data inspection, data analyses and model choice. The data deliberately has no context.

The assignment:

The assignment is to investigate and analyse the attached data (DS_case_study_1.txt.gz) so as to produce a model that can be used to predict the “response”.

The first column of DS_case_study_1.txt.gz is the binary variable called “response”, the 16,562 other columns are binary columns that can be used to predict the “response”. There are names for each of the columns that can be used for prediction, but those names are of no consequence.

Deliverables:

1. A write up of what you did and why.
2. The code you used to perform the work (feel free to make the write up and code into a combined document).
3. A description of model that can be used to predict the “response”, with a description of how well the model does (in terms of prediction the response).

Expectations:

There is no such thing as a perfect model, only what can be done is certain amount of time. We are not expecting you to work on this task for more than 4 hours. We value looking into the write up of your process and your thinking around the problem as well as how well the model that you report on does in terms of prediction.