

1. How long did you take to complete the test?

I stucked to the limit amount of time, which was 2 hours.

2. Was it easy or challenging? Which parts of the test were easy and which were challenging?

I think the main challenge was the lack of time and I took some time in understanding the objective of the task.

3. What resources did you use to learn how to solve the test (i.e. Google, forums, books)?

My own personal machine learning and data science projects as well as a few Google searches (mainly stackoverflow).

4. Briefly describe the process that you went through to find the solution for the problem.

Firstly, I took some time planning the steps to approach the problem. Then the process I followed was:

1. Loading the necessary Python packages.
2. Preliminary data overview
3. Data cleaning: check missing or duplicated data
4. Preliminary data selection
5. Define the target, training and test sets
6. Data pre-processing: scale the numerical features and encode the categorical features
7. Model selection and run
8. Make predictions
9. (Evaluation of the model: not included due to time limitations)

5. Are there any other key experiences/notes that you would like us to know in regards to your experience in taking the test?

I think I spent a large amount of time thinking on the planning. I would have done this differently considering I did not included some of the most advanced steps I planned due to the time limitations:

- A more sophisticated preliminary data overview including visualisations
- A detailed feature selection based on:
 1. Study of the influence of the features on the target by means of techniques such as mutual information regression
 2. Study of the correlation between features to detect redundancy
- Better target definition on the code
- Selection of different models and motivation
- Evaluation of the models' performance throughout hyperparametric studies (e.g. regulation parameters), cross-validation and learning curves.
- Assessment and visualisation of the models' predictions on the test set.

Unfortunately, I had to remove un-finished code covering some of the points mentioned above. Furthermore, I would have liked to better explain and motivate each step I included in the code.

If I could have started over, I would not have spent so much time on the planning. My aim was to follow the best approach I could provide, but I underestimated the amount of time I would need to include everything I intended to.