

Take-home Assessment

1 Description

This final practical is based on the material covered in lectures and previous practicals. You should write a practical report that should include a description and evaluation of the work done of not more than 2500 words excluding tables, graphs and images. The final practical will contribute 80% of the final mark. The deadline for uploading completed reports as a PDF to Moodle is **Friday 4th December 2020, 5pm**.

Additionally, you will need to submit your code as Jupyter notebook(s) to the Moodle webpage. Assessors may run your code, but you will **not** be assessed on the quality of code writing, **nor** will you be assessed on the basis of where your system's results rank amongst others. The assessment will be based on the report itself and on clarity of description of the work done, evaluation performed, and insights gained.

2 Dataset

The dataset is Clickstream data for online shopping

You should download the dataset from the UCI ML Repository and are encouraged to read: this paper
Note that the description of the data is also in the downloaded zip file.

3 Your task

Your task is to build a machine learning pipeline to predict the target value based on the other insightful variables in the dataset. The target value (price 2) represents whether a customer paid a premium price for the product selected. Imagine that you have been asked by an e-commerce website owner to build this model, perhaps because the owner wishes to dynamically alter the site based on the user or their behaviour to maximise profit.

Your implementation and report should include the following steps:

- *Data exploration*: note that the dataset contains a combination of categorical and numerical-valued features. Explore different features in the dataset, gain insights from the data, and report your findings. Consider whether it makes sense to exclude any of the features.
- *Machine learning algorithms implementation*: apply machine learning algorithms that you learned about in the previous practicals. Find out which algorithm works best and report your results.
- *Evaluation*: look into different ways and measures for evaluating the algorithms. Report your results and present your findings for the best-performing ML algorithm.
- *Visualisation and dimensionality reduction*: look into dimensionality reduction, and use visualisation as you think appropriate to reveal properties of the dataset and/or your model.