

Data Mining Assignment 2 Hotel Search

Yorick Mengelers¹[2557326], Ankur Anmol²[2701807], and Qiao Ren³[11828668]

¹ Vrije Universiteit, Amsterdam, the Netherlands y.mengelers@student.vu.nl

² Vrije Universiteit, Amsterdam, the Netherlands a.anmol@student.vu.nl

³ Univeristy of Amsterdam, Amsterdam, the Netherlands qiao.ren@student.uva.nl
group number: DMT-2021 160

1 Schedule

We made a schedule on when to do what and who did what. This is shown in table 1.

Table 1. Schedule of when we perform which tasks

Week	Tasks	Name
19-23 Apr	1.Read assignment, Divide tasks 2. Specify research question 3.Load dataset	All
26-30 Apr	1.Dataset exploration (plots, distributions, and pre-pre-processing) 2.Approach (regression, classification or ranking) 3.Previous approaches 4.Validation scheme and feature selection	Ankur Qiao Yorick Qiao Yorick
3-7 may	1.Finish pre-processing 2.Start building models	Ankur Qiao, Yorick
10-19 may	1.Finish building models 2.Make prediction on the test set 3.Evaluate model	Qiao, Yorick
15-21 may	Write report	All

2 Task Division

We divided the tasks in the following way:

Yorick: Yorick was responsible for data exploration, data visualization, data normalization, identifying outliers, dealing with missing data, building up three regression models, training the models, making prediction on the test set and writing the report.

Qiao: Qiao was responsible for reading previous works, feature engineering (correlation matrix, principle component analysis, dealing with imbalanced data set), building up three classification models, training models on subsampled dataset, implementing evaluation approach and writing the report.

Ankur: Ankur was responsible for data exploration, data visualization, feature selection, and preprocessing the data.

3 Reflection of the overall cooperation

We are willing to share our ideas, for example what model could be built to do the ranking, how to speed up the training process and how to evaluate the ranking result. During each meeting, we made a plan on what to do, when to finish it and we divided the tasks. Each of us gave an update on what we have done during the meetings. We gave feedback to each other on how good his or her work is and what could be improved. Our team members are willing to help each other when any of us has questions. We respond quickly when anyone ask questions in the group. Our discussions with TAs are also very helpful.

What could be improved is time management. It is better to start early on building up the model. Because it is out of our expectation that the training and prediction took a large amount of time. Besides, even though we made a plan in the beginning with all the milestones, we did not strictly follow the plan. Our actual process is a bit slower than our plan. To improve on it, we think we need more selfdisciplines to execute the plan. Finally, a small tip is that it is better to have more discussions with TA. So our questions could be solved in time.