Process Report

Group 144: Daniel Bemerguy[2699826] and Corien Westveer[2714024]

Vrije Universiteit Amsterdam

1 Project schedule

First we will describe our project schedule in table 1.

Table 1. Project schedule

Date	Task
April 23	First meeting to discuss how we want to approach this assignment
April 23 - April 30	Everyone looks for answers for task 1 and task 2
April 30	Meeting to discuss our answers for task 1 and task 2
April 30 - May 3	Looking for ideas for task 3, writing report about task 1
May 3	Meeting where we discussed our ideas for task 3
May 3 - May 7	Programming the code for task 3, writing report about task 2
May 7	Meeting to discuss how we want to approach task 4
May 7 - May 17	Creating machine learning models (task 4)
May 17	Meeting to discuss our approaches in task 4
May 17 - 20	Trying some changes in the models
May 21 - May 22	Writing the assignment and process report

2 Who contributed to what task

For task 1 everyone reads papers of solutions of the 2013 Kaggle challenge. We shared some of the ideas we found in this papers during a meeting and Daniel wrote the report section from this task. Emmanuel was sick the week we did this task, and therefore did not contribute to this task. Both the coding part and the report writing of task 2 was done by Corien. For task 3 everyone looked into some options for the feature transformation in the literature, in a meeting we decided what features we wanted to create, how we wanted to handle the missing data etc. Daniel did the coding of the feature transformations and wrote the report section about it. For task 4 we made our models independently. Corien made a random forest model and Daniel a made a Extreme Gradient Boosting Classifier. At May 17 Emmanuel decided to quit the course, since he did felt that he could not spend time to work properly on this assignment due to personal circumstances and he did not want to take credit for work he did not do.

3 Reflection on overall cooperation within the team

Generally spoken the cooperation between Daniel and Corien went well. The work was more or less equally divided. There was a willingness to help each other with problems and to questions asked in the chat came response within reasonable time. Also we made a planning on what task we wanted to do in which week before, which helped in not needing to do everything at the last minute.

One thing to reflect on is that one of the group members decided to quit at May 17th, as described in the "Who did what" section. This decision was made solely by himself, we did not ask him to do that. Already from the beginning there where a lot of signals that he did not contribute equally to the project as Daniel and Corien did, but we never said anything about that. When looking back, this was not a good solution. It would have been way more transparent if we would have had a conversation about the unequal contribution earlier during the project.

Another thing that was not the best way of coorperating was that we split up the task of building a model completely, we worked individually on a different model. But what we noticed was that we both had very different problems, which could have been solved earlier when we would have cooperated together on one model. For example, Daniel struggled with how to order the data, because he did predict a binary value (so for example either a 0 or 1 for *click_bool*) instead of a likelyness of beining clicked/booked. Corien already predicted the likelyness from the beginning, so this problem would have been solved earlier when we would have cooperated on one model. Another example of this is that Corien was struggling with that there were multiple target variables (click and booking). Daniel found for this the solution of a multi-output method, which could have solved Corien's problem earlier when we would have been cooperating more. So the conclusion is that we could have created a better models if we would have cooperated more on one code, in that case we could have used both our strengths to come up with the best solution.