

Assignment 2 Data Mining Techniques

Christos Petalotis^{1[2739394]}, Julia Wesselman^{2[2715996]}, and Matteo De Rizzo^{3[2749303]}

Vrije Universiteit Amsterdam, 1081 HV Amsterdam, Netherlands

`c.petalotis@student.vu.nl`

`m.de.rizzo@student.vu.nl`

`j.a.wesselman@student.vu.nl`

1 Process report

In order to complete the project, the team met up together via Zoom on ten different occasions. Certain planning decisions were taken during those meetings, and each group member would work on them on the time until the next meeting.

Here is reported the list of meetings that took place and a brief summary of their content:

- 25th April Discussion of a plan for the project. This would be to individually get familiar with the data set and study the Business Understanding section to be able to start working together on it during the next meeting.
- 29th April As we agreed that business understanding would be performed individually, only a discussion about the findings was necessary. Julia would perform the task of writing the business understanding section on the report while Matteo and Christos would review the text, providing feedback and adding certain details that they deemed useful.
- 3rd May Julia and Christos would begin working on the Exploratory Data Analysis, while also placing the findings in the report. The report would then be reviewed by the whole team.
- 8th May Matteo and Julia would work on the following steps: coming up with novel attributes and start to prepare the data set for model analysis. The report would also be updated throughout the process.
- 12th May Matteo, Julia and Christos would finish up adding previously decided novel attributes and clean the data set.
- 14th May The team decided to take a small break from the project because of academic and personal reasons.
- 23th May The EDA still required some work. Christos was assigned with finishing this task, while Matteo and Julia would look up information regarding fitting models.

- 25th May While Christos and Julia worked on setting up the model, Matteo would run multiple versions of modified data sets to estimate which would provide the highest score.
- 26th May The model would be run multiple times by Matteo, adjusted each time with new variables, with Christos and Julias's help
- 27th May The team met to discuss the last details and finalize the report, especially the "Modelling and Evaluating" and "Process Report" Sections. The final submission on the Kaggle competition were made during this meeting.

When planning for the meetings, the team decided to opt for having recurrent Zoom sessions, rather than meeting in person at the faculty. The reason for this was the discrepancy between each other's time schedules, as well as the frequent recurrence of meetings. Outside of planned meetings, it was challenging to be working at the same time, because each team member had a different preferred working time, as well as other planned activities. Nonetheless, everybody managed to accomplish the tasks that were assigned to them during the meetings on time, and when some difficulties arose, there were conversations about these on a WhatsApp group, where everyone immediately tried to aid in the task.

In general, the members of the team managed to work with each other very well, having no communication issues. Furthermore, each member provided assistance when necessary, either on the code level or to explain certain concepts that might not have been clear at first.

Regardless of different daily schedules, almost for the entire duration of the span of the project, everyone did their best to be available when needed.

Lastly, while the skills of different members would vary, each one contributed to the project as much as possible, working on both parts that were comfortable and challenging, with the help of others if necessary. There were some knowledge differences between each member of the group, with Christos and Julia being more familiar with Python concepts (which was used to write the code for both assignments), while Matteo was more knowledgeable in terms of ML and AI concepts. Nevertheless, all sides worked hard to get improve on their "weak" abilities, which turned out to be quite a valuable and pleasant collaboration.