

# Data Mining Techniques 2, Process Report - Group 29

Caterina Buranelli (2718588), Ignas Krikštaponis (2718072), and Charlotte Felius (2648506)

Vrije Universiteit Amsterdam  
<https://www.vu.nl/nl/index.aspx>

## 1 Schedule and Roles

Table 1: Project timeline.

Date	Caterina	Ignas	Lotte
19/4 - 01/5	Taking DataCamp Machine Learning Fundamentals courses.		
02-7/5	Holiday		
8-12/5	EDA	EDA	Related work
12-13/5	Feature engineering	Feature engineering	Modelling
14-15/5	Modelling and first Kaggle submission		
16/5	Parameter tuning	Parameter tuning	RF and sampling
17/5	Feature engineering	Feature engineering	Testing RF
18/5	Report	Feature engineering	Report
19/5	Report	Parameter tuning	Report
20-21/5	Finishing the report and the process report		

## 2 Reflection

We were all beginners in machine learning and therefore we decided to start by taking fundamental courses on ML at the beginning of the second assignment. After we finished those courses and the spring break, we started working on the assignment simultaneously.

We used WhatsApp as the primary means of communication during this project, in order to quickly exchange opinions, send files and useful links. We often video called not only with the purpose to discuss some topics, but also for motivation and working together, as if we were in a physical meeting. Our first intention was to meet physically but we had to switch to online meetings only because one of our members got infected by COVID-19.

WhatsApp enables its users to extract conversation data which we used to "mine" the messages sent by group members. In Figure 1 we can see that all members actively participated in discussions. Most of the communication took place after May 10th with most messages sent on the 14th.

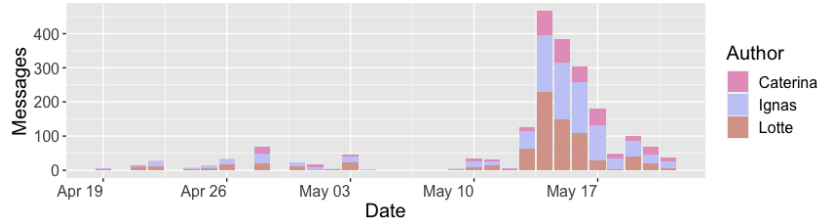


Fig. 1: WhatsApp messages.

The mean time gap between messages among different members - response time - (excluding time intervals larger than 4 hours to account for nighttime) was found to be around 5.5 minutes.

We shared the code on a repository in GitHub. On Figure 2, the number of commits per team member is plotted.

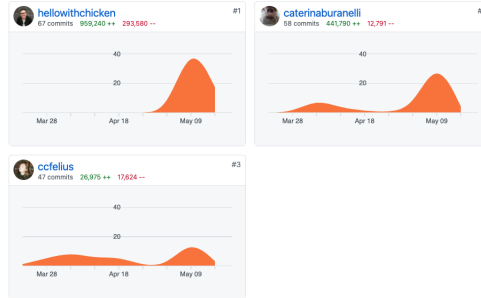


Fig. 2: Github commits.

All the members of the group were very approachable and quickly (as the mean response time highlights) ready to help each other when there were some doubts or issues with the code. Meetings were handled in a very *agile* way. We did not register differences on the amount of work each one did.